One-Dimensional Lattice Gasses with Soft Interaction

Application to Nucleosome Positioning in Yeast

Brendan Osberg

München 2015

One-Dimensional Lattice Gasses with Soft Interaction

Application to Nucleosome Positioning in Yeast

Brendan Osberg

Doktorarbeit
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Brendan Osberg
aus Halifax, N.S. Kanada

München, den 19. August 2015

Erstgutachter: Prof. Dr. U. Gerland
Zweitgutachter: PD Dr. P. Korber
Datum der mündlichen Prüfung/Date of oral examination: 25.03.2015

Contents

Zussamenfassung

Eukaryotische DNA muss mehrere Stufen einer organisierten Kompaktifizierung durchlaufen, um in die räumlichen Grenzen eines Zellkerns zu passen. Die erste Stufe dieser Kompaktifizierung beinhaltet den Aufbau von Nukleosomen durch die Verbindung von DNA und Histon-Oktameren. Die Anordnung dieser Nukleosome entlang der DNA hat wichtige Einflüsse auf die Organisation höherer Kompaktifizierungsstufen. Zusätzlich zu ihrer strukturellen Funktion hat die Positionierung von Nukleosomen entlang eines Genoms, sowie die Wechselwirkung von Nukleosomen untereinander wichtige Implikationen für die Regulation von Genen. Dicht gepackte Nukleosome neigen dazu, Promotorregionen von der Transkription auszuschließen, während eine lockere Packung von Nukleosomen in der Regel zur Hochregulation der entsprechenden Gene führt.

In dieser ersten Stufe kann die Positionierung von Nukleosomen effektiv als ein eindimensionales System beschrieben werden. Viele Faktoren tragen zur Positionierung von Nukleosomen entlang einer DNA bei. Hierzu zählen die Nukleotidsequenz der DNA, aktive "Chromatin remodellers", sowie der Wettbewerb um Bindungsstellen zwischen Nukleosomen untereinander und mit anderen Bindungsproteinen. Die Einordung der einzelnen Faktoren ist zentraler Bestandteil dieser Arbeit, wobei Hefe als Modellorganismus dient. Im Verlauf der Arbeit werden allgemeine physikalische Fragen hinsichtlich der Kinetik eindimensionaler Adsorptions- und Desorptions-Prozesse aufgeworfen. Das übergreifende Ziel ist daher die Errichtung einer Brücke zwischen Daten getriebener biophysikalischer Forschung und statistischer Physik. Infolgedessen besteht diese Dissertation aus fünf verschiedenen, jedoch untereinander verwandten Projekten.

Diese Arbeit beginnt mit einem Überblick über Hintergründe und einführende Beobachtungen in Kapitel 1. In Kapitel 2 liegt der Fokus auf den Gleichgewichtseigenschaften der Nukleosom-Positionierung. Experimentelle Nukleosom-Daten von einem Dutzend verschiedener Hefearten werden verwendet, um die Anordnung von Nukleosomen in der Nähe einer Barriere zu modellieren. Bei der Barriere handelt es sich um das stark positionierte +1 Nukleosom mit geringstem Abstand zur Transkriptionsstartstelle (abwärts).

Es wird gezeigt, dass die Berücksichtigung von "Weichheit" der Nukleosomen, aufgrund von bekannten biophysikalischen Effekten, eine einheitliche Modellierung der Nukleosom-Positionierung ermöglicht. Da die Struktur von Nukleosomen verhältnismäßig konsistent zwischen verschiedenen Arten ist, ist das beschriebene Modell sowohl minimalistisch als auch physikalisch sinnvoll. Angefügt an das Kapitel ist eine veröffentlichte Arbeit über die Anordnung von Nukleosomen in einem Dutzend verschiedener Hefearten. Diese Verffentlichung basiert auf der statistischen Physik des Gleichgewichts, sowie auf numerischen Monte-Carlo-Methoden zur Berücksichtigung von aktiven Prozessen.

Obwohl DNA-Bindungspositionen durch Histone dominiert sind, gibt es wichtige Orte an denen eine Bindung von anderen Proteinen, wie zum Beispiel von Transkriptionsfaktoren, möglich ist. Diese Faktoren dienen zur Regulation der Transkription und beeinflussen die Anordnung der Nukleosome. In Kapitel 3 betrachten wir die Wechselwirkung kleiner Transkriptionsfaktoren, die mit spezifischen Loci binden und die Positionierung der benachbarten Nukleosomen verschieben kann. Ein Dominoeffekt auf andere benachbarte Nukleosomen wird ebenfalls beobachtet. Eine solche Verschiebung in der Anordnung von Nukleosomen induziert

eine Kooperativität zwischen Transkriptionsfaktoren, deren Reichweite mehrere Nukleosome umfassen kann.

Im Kapitel 4 wird die Rolle der genetischen Sequenz auf die Nukleosom-Positionierung näher betrachtet, die schon vielfach Gegenstand der Forschung war. Hier wird die genetische Sequenz als energetische Landschaft betrachtet und ein neuer Weg zur Bestimmung von Sequenz-Präferenzen aus Daten ber die Nukleosom-Positionierung dargelegt. Es wird gezeigt, dass die experimentell beobachteten Dichteverteilungen in Hefe, kombiniert mit der in Kapitel 2 hergeleiteten Interaktionsenergie zwischen benachbarten Nukleosomen, genutzt werden können, um die Sequenz-Präferenzen zu quantifizieren. Dieser Prozess wird jedoch durch den derzeitigen Mangel an spezifischen Daten über zwei-Körper-Korrelationen zwischen benachbarten Nukleosomen erschwert. Aus diesem Grund ist der "Amöben" Optimierungsalgorithmus auf die verfügbaren Daten angepasst, wie in Kapitel 4 beschrieben.

In Kapitel 5 wird der Fokus in Richtung der Dynamik des eindimensionalen Füllens verschoben. Es wird gezeigt, dass der kinetische Prozess der Gleichgewichtseinstellung durch eindimensionale reversible Adsorption qualitativ anders und sehr viel schneller ist, wenn weiche Interaktionen zwischen benachbarten Teilchen erlaubt sind. Es ist seit langem bekannt, dass die Adsorption von "hard rods" in einer Dimension ein "jamming" Phänomen verursacht, das nur durch sehr langsame, kollektive Umordnungsprozesse zu organisierten "Arrays" mit hoher Dichte gelöst werden kann. Mit der Einführung von weichen Wechselwirkungen im Nukleosom-Modell wird jamming durch eine neue Phase, die wir als "cramming" bezeichnen, umgangen; der Übergang ins Gleichgewicht erfolgt auf Zeitskalen, die um Größenordnungen kürzer sind. Dieses Prinzip wird hinsichtlich seiner Anwendung auf die Anordnung von Nukleosomen präsentiert. Die wichtigsten Erkenntnisse hierzu sind in der angehängten Publikation beinhaltet.

Abschließend wird die Dynamik des eindimensionalen Adsorption- und Desorptions-Problems von weichwechselwirkenden Teilchen in einer allgemeineren Weise betrachtet. Mit endlichen nachbarschaftlichen oder benachbarten Wechselwirkungen entsteht eine reichhaltige Dynamik, einschließlich eines seltsam, nicht-monotonen Dichteverlaufs in der Zeit. Die theoretischen Grundlagen dieses Effekts werden in einem Manuskript, das diesen Text schlussfolgernd abschließt, präsentiert.

Abstract

Eukaryotic DNA must undergo several levels of organized compaction in order to be packaged within the spatial confines of the cell nucleus. The first level of this packaging involves the formation of nucleosomes by wrapping DNA around histone-octamers. The arrangement of nucleosomes along the length of the DNA has important influences on the way higher levels of packaging are organized. In addition to this structural role, the positioning of nucleosomes along the genome –and in relation to one-another– has important implications for the regulation of genes. Tightly-spaced nucleosomes tend to occlude promoter regions from transcription machinery, while looser configurations tend to be up-regulated.

At this level, nucleosome positioning can be treated as an effective one-dimensional system. Within this framework, many factors contribute to the positioning of nucleosomes along the DNA: genetic sequence, active remodellers, and competition for binding sites with other binding proteins and with one-another all play a role. How to disentangle these effects is a central question that will be explored in this work using yeast as a model organism. In the process, however, more general physical questions will arise regarding the kinetics of one-dimensional adsorption/desorption processes. The over-arching goal is to provide a bridge from biophysical, data-driven work to more pure statistical physics; thus the work is comprised mainly of 5 somewhat separate, but related projects.

This thesis will begin with an overview of background information and introductory observations in Chapter 1 to provide context. Chapter 2 will then focus on equilibrium properties of nucleosome positioning. Experimental nucleosome data from a dozen different species of yeast will be used to model the pattern of nucleosome formation near a 'barrier' –in this case, the strongly positioned +1 nucleoseome nearest (downstream) to the transcription start site. It will be shown that accounting for 'softness' in nucleosomes, due to known biophysical effects, allows for a unified model of nucleosome positioning. Since nucleosomes are relatively structurally consistent across very different species, this represents a model that is both parsimonious and physically sound. The published work studying the nucleosome positioning patterns of a dozen species of yeast is included and relies on equilibrium statistical mechanics, as well as a Monte Carlo numeric scheme to account for active processes.

While histones clearly dominate the landscape of DNA binding positions, important loci admit binding by other proteins such as transcription factors which serve to regulate genetic transcription and influence nucleosomal patterning. In Chapter 3, we consider the interaction of small transcription factors which bind specifically to loci on the DNA and shift the positioning of the neighboring nucleosome, with a corresponding domino effect on other nucleosomes in the vicinity. Such shifts in nucleosome patterns can create nucleosome-mediated cooperativity between transcription factors, even when separated by intervening nucleosomes.

Next, in Chapter 4, we will consider the role of the genetic sequence in nucleosome positioning, an effect which has also been the subject of considerable research. We will refer to this as the energetic 'landscape' of the genome and present a new way of inferring this sequence-preference from nucleosome positioning data. We will see that the experimentally observed density patterns in yeast, together with the interaction-energy of neighboring nucleosomes that was derived in Chapter 2, can be used to quantify this sequence preference. This effort, however, is complicated by the lack of specific data characterizing the 2-body correlation

between neighboring nucleosomes. For this reason, the 'amoeba' optimization algorithm is adapted to fit the available data, as described in Chapter 4.

In Chapter 5, the focus will shift to the dynamics of one-dimensional filling. It will be shown that the kinetic process of equilibration through one-dimensional reversible adsorption is qualitatively different, and much faster, when one allows for soft-interaction of neighboring particles. It has long been known that 'hard rods' adsorbing randomly in 1 dimension undergo a jamming phenomenon which can only be resolved into densely packed arrays through very slow collective rearrangement processes. Upon introduction of softness to the nucleosome model, however, jamming is circumvented by a new phase we term 'cramming'; equilibration can then proceed orders of magnitude faster. This will be reviewed with specific application to the problem of nucleosome adsorption which has been of interest recently in light of new experimental work and the attached publication highlights the main findings.

Finally, the dynamics of one-dimensional adsorption-desorption with soft-interacting particles are considered in a more general way. With finite neighbor interactions, a rich new set of dynamics emerges, including a curious non-monotonic density trace in time. The theoretical underpinnings of this effect will be provided in the attached publication which concludes this text.

1 The Nucleosome

Within the nucleus of eukaryotic cells, DNA is compacted on multiple levels; the need for such condensed packaging can be readily understood even from just spatial considerations. The DNA of, say, human cells, if extended lengthwise, with chromosomes end to end would stretch over 1m. in length and yet is compacted (in a non-random way) within a cell nucleus microns in diameter –a remarkably illustrative example of the complexity hidden within living organisms.

At the most basic level of this condensation, the long string of DNA is wrapped just under two times around a histone-octamer core to form a nucleosome, the repeating unit of chromatin, while the DNA extending outward in either direction continues on to form adjacent nucleosomes or to bind with other factors[1]. At the next level of packaging, it has been found in some organisms that these nucleosomes are stacked alongside one-another into long fibers [2, 3]; notably, however, such fibers have not been found in the human genome[4], and the generalizability of such higher-order structure in chromatin continues to be a matter of ongoing research. Still further levels of organization are required to contort these long fibers into superhelical arcs to form the (relatively) massive contours of chromosomes that are more familiar in images of chromosomes from textbooks.

These latter descriptions of genomic architecture are more complex than what is necessary at the level of description employed throughout this work, though their importance for biological purposes should not be overlooked. At increasingly fine levels, they contribute to relevant regulatory functions. Heterochromatin, for example, is tightly packed with nucleosomes and is often characterized by the hypoacetylation and/or methylation of histones[5, 6]; euchromatin, on the other hand, is more loosely packaged and transcriptionally active, and is associated with converse patterns of post-translational modification. Epigenetic changes in features of chromatin structure such as these can even be preserved across multiple generations[7], and their simulated effects on gene expression have been studied[8].

Fig. 1.1 delineates the various size scales of chromatin organization, and provides some perspective to our model. In keeping with conventional notation, we use 'primary structure' to refer to nucleosome positioning, while possible fibers and higher organization are referred to as secondary and tertiary structures respectively. In practice, however, the properties of these structures are somewhat overlapping at various scales. For example, it has been argued that nucleosome 'stacking' at the secondary level exerts influence on positioning at the primary level [9].

Nevertheless, throughout this work, we make use of the Kornberg-Stryer model[10, 11] and assume that the higher-order steric and architectural features of chromatin can be omitted from a description of the positioning patterns of nucleosomes along the length of the DNA. This amounts to positing that a one-dimensional coordinate system along the length of the DNA suffices to capture the most salient properties of nucleosome positioning.

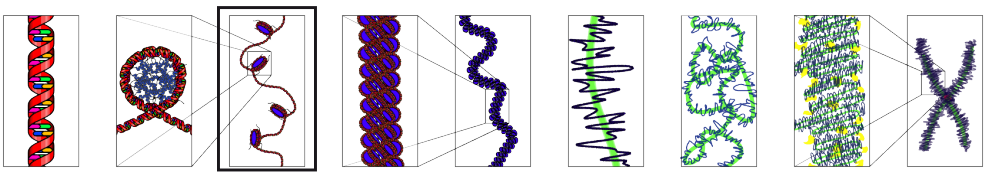


Figure 1.1: Many length scales exist between the description of chromatin as a double-helix of DNA and familiar microscope images of a chromosome. The circumscribed panel indicates the scale of our analysis, the so called 'beads on a string' picture of chromatin primary structure. Secondary 'fibers' are adjacent to the right and higher order organization follows, although these levels of configuration are beyond the scope of description here. Image adapted from the work of Richard Wheeler, from the wikimedia commons under the Creative Commons License.

Hence, we focus on the primary structure of chromatin, beginning with the nucleosome itself which is roughly 10 nm in diameter and is surrounded by 147 bp of DNA[12]. Generally, the nucleosome is comprised of 8 histone proteins: 2 each of H2A, H2B, H3 and H4 at its core; variants of these histones have been noted. It is typically formed in a two step process, starting with a tetramer of H3, H4 histones adsorbing to the DNA, followed by H2A, H2B histones[13]. An additional H1 'linker histone' at the interface with DNA is often also present[14]. An important point is that when two adjacent nucleosomes are close together, the binding of one histone octamer with DNA sterically prevents the other from doing the same; an illustration is provided in Fig. 1.2.

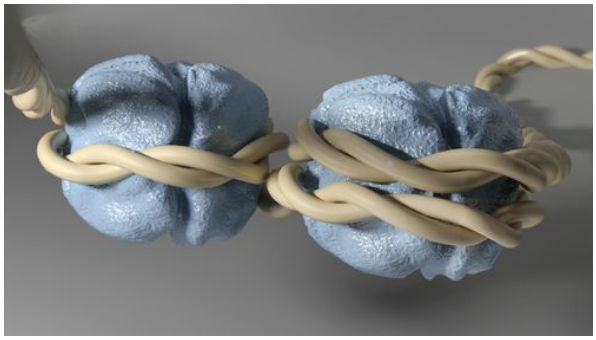


Figure 1.2: Illustration of two adjacent nucleosomes along DNA. At close proximity, the binding to intervening DNA can become 'competitive'. Image credit to Christoph Hohmann, Nanosystems Initiative Munich; used with permission.

The histone octamers depicted in Fig. 1.2 bind to the minor groove of the DNA, and require significant bending of the DNA molecule around the histone surface to keep the entire unit stable. Since multiple nucleosomes are prevented from adhering to the same stretch of DNA simultaneously, it is convenient to visualize DNA as a long substrate upon which proteins, such as histones, bind and reside as though they were 'beads on a string'. This convenient visual metaphor captures the much-simplified topology of this view. Despite the simplification, it has been shown that this picture of chromatin has strong predictive and analytic value since its development[11].

Within this single dimensional picture, the positioning of nucleosomes is determined by a

host of factors[14], including sequence preference and active remodelling agents. In addition to these, a further influence on positioning is the competitive binding of histones with other factors, and with one-another; the cumulative effect of this competition between many binding factors results in statistical positioning, where nucleosomes are 'stacked' closely together. Analyzing this behavior, as well as the individual interactions between neighbor nucleosomes will form the bulk of the following chapter. For the moment, it is useful to review the analysis of available experimental data that will serve as the point of reference to evaluate the success of our models.

1.1 Available Experimental Data

The majority of nucleosome positioning data used as a reference throughout this work is gathered through the use of micrococcal nuclease, or MNase, applied to the yeast genome –a standard model organism that has been well studied. Theoretical analysis presented here has been made possible through availability of data from experimentalists in the field [15, 16, 17, 18], which we briefly review the processing of here. Other techniques have also been developed for determining nucleosome positioning, with qualitatively similar results (albeit with more specific resolution)[19]. A concise summary of the available positioning data is provided by Bai and Morosov:

> "The detailed methods used for nucleosome mapping can vary slightly in different labs and for different species but the principle remains the same [...]. Live cells or spheroplasts (sometimes crosslinked with formaldehyde) are made permeable and treated with MNase, an endo-exonuclease from Staphylococcus aureus that preferentially digests linker DNA (uncovered DNA between neighboring nucleosomes) versus nucleosomal DNA. The reaction is usually carried out to the extent that most of the chromatin is digested to mononucleosomes, with subpopulations of di- and trinucleosomes. The digested chromatin is sometimes purified further by immunoprecipitation with histone antibodies before the crosslink is reversed (if necessary) and the DNA segments are extracted. Mononucleosomal-sized DNA fragments (150-200 bp) are selected by gel purification, and their locations on the genome are mapped by either hybridizing to DNA microarrays or high-throughput sequencing followed by the alignment of sequence tags to the reference genome."[20]

One point must be added to the above description: in order to obtain positioning data centered to the midpoint, or dyad, of the nucleosome, the sequence matches obtained from the 5' ends of the mononucleosomal-sized DNA fragments are generally shifted in the 3' direction by an offset of half of the average length of the segments. As described in the supplementary materials of one of these experiments: "All tag 5' coordinates were shifted by 73 bp in the 3' direction to identify the putative nucleosome dyads."[16]. Histograms of data used in this text are then intended to represent the distribution nucleosome centers.

It is also worth noting that a similar technique has also been combined with a "modified protocol" to detect DNA fragments smaller than mononucleosome fragments, and can serve to indicate positions of partial nucleosomes, or transcription factor binding[21], the effect of which will be considered in Chapter 3.

There are some limitations to the information that can be gleaned from these experiments. First, since the number of cells used in the experiment cannot be known exactly, available read data is scaled by an unknown normalization factor α corresponding roughly to the number of cells in the experiment. It is reasonable, however, to assume a single, global constant α across the entire genome, which preserves the validity of *relative* comparisons of nucleosome density between one region and another within a single experiment.

Secondly, during this process, information regarding which mononucleosomes had originally been adjacent to one-another on the chromosomes is lost, and only the sum of individual positioning data from many separate cells is preserved (although later techniques[19] are able to partially preserve this information).

Nevertheless, MNase experiments provide an invaluable description of nucleosome positioning along the genome. The natural question then arises: what regions of the genome are enriched with nucleosomes, and what determines this positioning? To begin addressing this question, we consider the degree to which nucleosomes are organized into characteristic positions at all. Taking positioning data from each of the genome's two strands[1] we quantify the correlation in read matches between these two halves of the experimental data, using a scalar overlap function $O(x)$ for each chromosome

$$O(x) = \frac{1}{N} \times \left(\sum_i \text{Watson}(i) \times \text{Crick}(i + x) \right), \tag{1.1}$$

as a function of the offset x between them. In Eq. 1.1, the summation index i is taken across the entire length of the chromosome, and the normalization factor N is defined as the maximum value of the overlap.

Eq. 1.1 represents the degree of consistency found in positioning data between one half of the nucleosome positioning data (the Watson strand, in this example), and the other half (the Crick strand, likewise), over many cells. Assuming random cell-to-cell positioning[2], a plot of Eq. 1.1 vs. x should amount to simple noise independent of x. However, Fig. 1.3 using available experimental data[16] shows quite prominent features.

Of course, it has long been known nucleosome positions are not entirely random from one cell to another. There are many factors contributing to the positioning patterns of nucleosomes in yeast and other organisms, as can be seen in nucleosome positioning maps[22]. A summary of these effects is outlined in Ref. [23], and a major goal of this thesis is to ascertain the relative contributions of each of them. Thus, while Fig. 1.3 makes no novel claims, it nevertheless contains several illuminating features for discussion.

First, the very sharp peak at precisely $x = 0$ is indicative of nucleosome positions that are defined specifically at the base-pair level. The somewhat broader peak behavior over the range $x \pm 50$ illustrates that positioning data is also subject to a certain amount of statistical fluctuation –the positioning of some nucleosomes is rather 'fuzzy'[15]. Secondly, the recurring features at $\pm$ 165 bp underscore the periodicity of nucleosome positioning. The

[1] assuming either of the two 5' ends of the extracted mononucleosome strands are equally likely to have been selected for sequencing

[2] due to the large number of cells used in such experiments, auto-correlated strands originating from the same nucleosome *in the same cell* contribute negligibly).

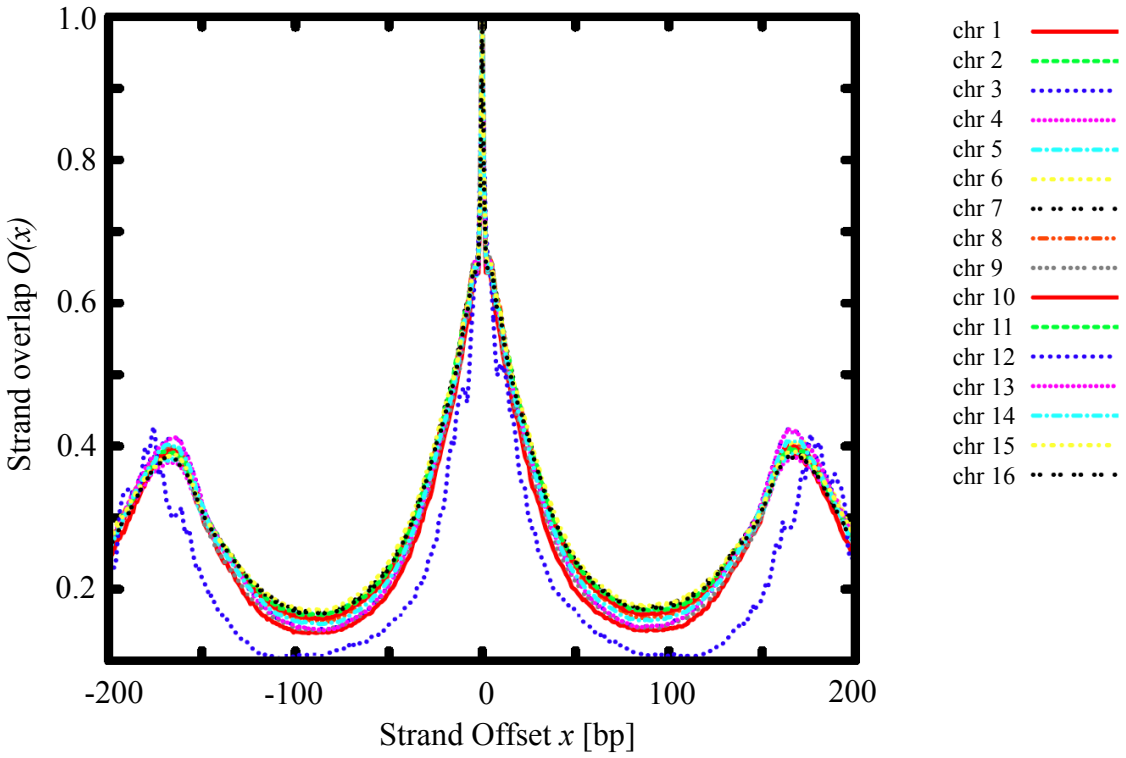


Figure 1.3: Scalar overlap of the nucleosome positioning data from Watson and Crick strands over the length of each chromosome as tabulated to the right of the figure. With the Watson strand oriented x bp offset from the Crick strand, the scalar overlap measures the consistency of positioning data obtained from the two strands as a function of x. In addition to the prominent maximum at $x = 0$, adjacent maxima can be observed at $x \approx \pm 165$, suggesting a trend in spacing between neighboring nucleosomes.

peaks produced by these overlapping read counts are due to correlation, not with the same nucleosome on the two strands, but with the one adjacent. In Chapter 2 it will be shown how occlusion of adjacent nucleosomes leads to extended patterns in nucleosome density due to statistical positioning, and thus, characteristic periodicity in positioning patterns. Finally, note that there is little difference in Fig. 1.3 between chromosomes in the shape of the curve of $O(x)$. One of the findings of Ref. [16] related to the grouping of genes into various categories according to the nucleosome density profile in the promoter region. In the course of this work, the distribution of these gene groups by chromosome was studied, showing no significant correlation; it seems that yeast genes are distributed, both by function and by chromatin structure, more or less randomly across the genome.

Hence, throughout the rest of this text, we treat all genes on an equal footing, making no further distinction with regard to the position or chromosome of each gene, and attempt to elucidate the mechanisms behind observed 1-D positioning patterns along the length of the DNA. It is, however, noted that there is heterogeneity in the properties of nucleosome positioning within each gene. For example, near the transcription start site (TSS) of most genes, there is found a characteristic nucleosome free region (NFR)[24] where the genome tends to be highly

depleted of nucleosomes, flanked by nucleosomes enriched with the H2A.Z modification[25] upstream and downstream –the latter of which tends to be strongly positioned and is referred to as the +1 nucleosome for the gene. In addition to being well positioned[26], +1 nucleosomes serve a particularly important regulatory role[27, 28, 29, 30], and it is especially in this region where nucleosome occupation changes most markedly under changing cell conditions [31]. Suffice it to say that the positioning pattern of nucleosomes near gene promoters has important biological implications, though the effects on transcription dynamics will only be discussed very briefly in Chapter 3. For the moment, the +1 nucleosome can be treated as an effective barrier to downstream nucleosomes whose locations, on a gene-averaged level, tend to be consistent with statistical positioning. These downstream patterns are the focus of the following chapter.

2 The Equilibrium Nucleosome Gas Model

As discussed in the previous chapter, a convenient simplification for the positioning pattern of histones on DNA is a 1-D lattice of base pairs along the length of the DNA, neglecting the effects of higher-order chromatin structure and superhelicity. Segments bound to histones can then be seen as 'rods' along the 1-D lattice. Within this framework, specifically positioned proteins such as transcription factors (TFs) or +1 nucleosomes downstream from a gene's NFR can serve as a boundary against which characteristic oscillatory patterns can be observed in the density of nucleosome positioning. One-dimensional models have been proposed as a means to rationalize these patterns on the basis of statistical positioning[11]. It has been argued that this effect, in conjunction with particularly well positioned histones and specifically-bound proteins, contribute significantly to nucleosome positioning[10].

Recently, statistical positioning models have been successfully applied to gene-averaged nucleosome data in yeast relative to the +1 nucleosome 'barrier' downstream from the genetic transcription start site [32]. Statistical positioning can be thought of as a link between the probability that a given position is occupied by a nucleosome and the occupation 'state' of other loci in the vicinity. In that respect, the problem is somewhat analogous to a well-studied and archetypal problem in statistical physics.

The Ising model was first studied by Wilhelm Lenz and Ernst Ising[33], the latter of whom it takes its name from. It continues to be employed as a model system in various applications in statistical physics and the key to its success lies partly in its ability to predict macroscopic, emergent behavior by extrapolation from simplified elementary interactions. While some of the more striking predictions of the model (e.g. phase transitions) are unique to higher-dimensional systems, the earliest work in this area is particularly relevant: the calculation of equilibrium properties of a set of magnetic spins lying in a one-dimensional array.

Fig. 2.1 depicts two such adjacent spins along a lattice interacting with each other and with the environment with characteristic energetics J and H respectively (assuming only nearest-neighbor interactions). Based on the strength of these interactions, amid thermal effects, the equilibrium probability of either spin-up or spin-down states at various positions can be calculated exactly. The key lies in taking statistical sums of the set of states available to each site throughout the chain. These calculations have long since been solved[34], and their biophysical application to DNA has already been made[35], but the imagery is a useful starting point as we explore some necessary methodology in greater detail in the next section.

2.1 The Transfer Matrix

The immediate relevance of this to nucleosomes may not be apparent until one considers that DNA is, in fact, a discrete lattice of nucleotides to which different factors can chemically bind;

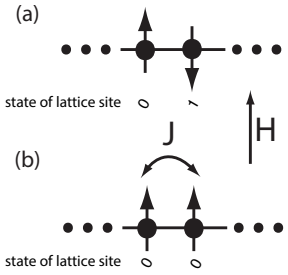


Figure 2.1: Minimal representation of the spin-Ising interaction to introduce the concept sites interacting with characteristic energetics depending on their states. In (a) spins are anti-parallel with only a single spin aligned with the local magnetic field. In (b), however, both spins are in 'state 0' and interact favorably with the local field H, as well as with each-other J. Thus, the arrangement in (b) is more likely than (a).

such binding can then be used to define the states of the lattice sites themselves. An obvious difference here is that the Ising spin lattice required only 2 unique states for each lattice spin. An array containing particles of size a lattice sites, however, will require $a+1$ unique 'states' to describe each possible position along the length of the particle (in addition to the state of being unoccupied). A schematic of one possible set of such states is given in Fig. 2.2.

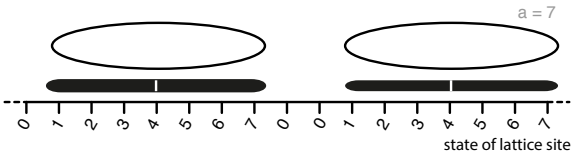


Figure 2.2: Schematic illustrating the 'states' of lattice site occupation by extended particles and the notation used for Eq. 2.1, for a footprint length of $a = 7$. Partially adapted from SI of Ref. [36].

Fig. 2.2 illustrates a set of states for lattice sites defined in sequence along the 'footprint' of the DNA where a binding agent is found. Assuming these particles represent nucleosomes, and assuming, for the moment, that nucleosome footprints cannot overlap, one possible matrix for such a 'hard-core nucleosome gas' (HaNG) is provided in Eq. 2.1,

$$T_{\mathrm{HaNG}} = \begin{pmatrix} 1 & e^{\mu} & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ 0 & 0 & 0 & 1 & & 0 \\ \vdots & & & & \ddots & 0 \\ 0 & 0 & 0 & 0 & & 1 \\ 1 & e^{\mu} & 0 & 0 & \dots & 0 \end{pmatrix} \tag{2.1}$$

where the entries adjacent to the main diagonal are the result of a sequential rule (e.g. a lattice site in state 3 *must* be followed by a lattice site in state 4). μ characterizes the energetic drive of the environment to form nucleosomes, and depends on the local environment. Note that that the matrix presented in Eq. 2.1 is degenerate for the sake of clarity: a single state could

be neglected without loss of accuracy, although with rank $a+1$, the mapping to all $a+1$ states is more intuitive. The transfer matrix given in 2.1 is not defined uniquely; other formulations are possible, such as those provided by Teif[35], or Chen[37] and produce identical observables. For a periodic array of length L, T_{HaNG} suffices to determine the partition sum

$$Z_{\mathrm{HaNG}} = \mathrm{Tr}\left\{T_{\mathrm{HaNG}}^{L}\right\} \tag{2.2}$$

which represents the sum of statistical weights of all possible configurations of the lattice, and can in turn be used to determine the probabilities of particular sites assuming given states. The assumption of a periodic lattice will prove to be immaterial in the limit $L \to \infty$. An alternative transfer matrix will presented later in this chapter; to avoid repetition, explicit calculation of state probabilities will be presented only once in Section 2.3 below. It is worth noting, however, that these probability calculations, using Eq. 2.1 produce the same two-particle correlation functions that have been developed previously[38, 39, 32]

$$\rho(x|0) = \sum_{k=1}^{\infty} \frac{\left(\frac{x}{a}-k\right)^{(k-1)} \Theta\left(\frac{x}{a}-k\right)}{a\Gamma(k)} \left(\frac{\overline{\rho}a}{1-a\overline{\rho}}\right)^{k} \mathrm{e}^{-\left[\frac{\frac{x}{a}-k}{\frac{1}{a\overline{\rho}}-1}\right]}, \tag{2.3}$$

where $\rho(x|0)$ is the probability of observing a particle at x given that one is fixed at 0, $\Gamma(n) = (n-1)!$, and $\overline{\rho}$ is the average density far from the barrier. In Eq. 2.3, the sum over k is intended to represent the sequence of nearest neighbor particles, next-nearest neighbor, ...etc. producing identical results to what will be shown below based on Eq. 2.1.

All of this, however, is predicated on the assumption of absolutely stiff particle boundaries. As we will see in the following section, further scrutiny of the binding behavior between histones and DNA demands we revisit this static characterization of the nucleosome footprint.

2.2 Finite Neighbor Interactions Between Nucleosomes

The model of nucleosomes as sterically exclusive segments of 147 bp is a useful abstraction in many ways and has demonstrative merit in reproducing experimentally measured nucleosome distribution patterns near a barrier[32]. One important way in which the above description of nucleosomes fails to reproduce the real biophysics of chromatin, however, is that nucleosomes do not conform so strictly to this picture of hard 147bp DNA segments for two reasons: First, nucleosomes are formed in a step-wise manner, incorporating H3 and H4 histone pairs first, followed by the H2A and the H2B histone pairs. In the interim, the histone tetramer occupies significantly less space on the DNA which may be significant during chromatin formation/reconstitution as we will see in Chapter 5. Secondly, even when nucleosomes are fully formed, they undergo transient unwrapping at their boundaries[40, 41], and allow for partial overlap[42, 43]. Unwrapping/rewrapping rates are fast on the time scale of overall adsorption/desorption rates[44], and produce measurable equilibrium probabilities of site exposure near the periphery of the nucleosomal 'footprints'. Towards the interior, however, it has been observed that "DNA unwrapping rate decreases dramatically with distance inside the nucleosome."[45]. It has also been suggested that such nucleosome 'breathing' is enhanced by the histone chaperone FACT[46] through the reorganization of the histone-DNA contacts near the boundary.

Even under thermal influences, segments of DNA will peel away, reversibly, from the histone surface from the outside in, before reattaching; since the bound state tends to be energetically preferable, reattachment occurs more rapidly than the reverse reaction, and the dynamics of these forward and reverse reactions are sufficiently fast to be in effective equilibrium. This transient unwrapping has been shown experimentally using FRET signals with donor-acceptor pairs attached to the DNA and histone core respectively [42, 41]. Furthermore, the temporary exposure of genetic loci via transient unwrapping has been shown to facilitate the binding of TFs[47, 45] at positions that would otherwise be occluded by a 'hard rod'-type nucleosome footprint –however with decreasing exposure at positions deep within the nucleosome. Thus, it seems clear that the soft-edge of the nucleosome footprint has real physical significance that ought to be included in a full model of chromatin primary structure.

2.2.1 The SoNG Model

In light of these observations, it has been noted[48] that the assumption made earlier of histones occupying exactly a base-pairs, with hard-exclusive interactions must be relaxed in order to present a more faithful description of nucleosome interaction with neighboring proteins. We assume, instead, that there exists only a central core of length n bp at the dyad of the nucleosome in addition to w bp in either direction that are preferentially bound to the histone surface, but which can, transiently, detach. A revised state-notation schematic is provided in Fig. 2.3

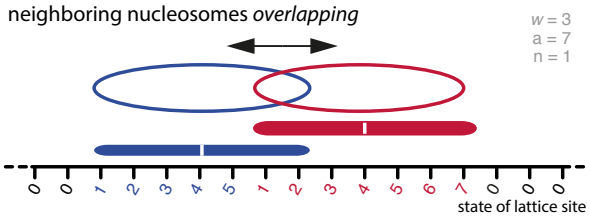


Figure 2.3: The 'states' of lattice site occupation by extended particles, again with $a = 7$, $w = 3$, now including the possibility of overlap between adjacent particles. The state definition corresponding to the second particle (i.e. to the right) supersedes the previous, and the left-most position of the particle is set as the 'reference' position. Partially adapted from SI of Ref. [36].

Due to the fast time scales of wrapping/unwrapping, this energetic cost of unwrapping can be treated as an effective soft repulsive potential for neighboring particles; resulting calculations are then similar to, e.g. Ref. [37]. There is, of course, no such 'special' segment within the nucleosome, and all DNA is subject to detachment. Rather, deep within the footprint of the nucleosome, DNA becomes increasingly *unlikely* to be removed from the histone surface and any unwrapping event that penetrates into segments approaching the dyad core would serve to destabilize the nucleosome entirely. For the sake of generality, we take a minimal hard core dyad length of $n = 1$ corresponding to the precise center of the nucleosome; thus $a = 2w + 1$. Typical unwrapping lengths generally do not extend beyond 40-50 base pairs inward from the outer edge, so this assumption is quite conservative.

An appropriate model of transient unwrapping must then reflect this incremental 'stiffness' approaching the center of the nucleosome footprint. A natural way to do this is to assume a constant energetic binding strength, ε, per unit length of DNA bound to the histone surface. The energetic favorability of the nucleosome then increases in proportion to the amount of DNA included. Indeed, this intuitive picture is supported by progressive increase in site exposure towards the outer edge of the nucleosome. The corollary of this assumption is that if two adjacent nucleosomes, separated by Δx bp both admit the possibility of partial unwrapping, then there exist states in which $\Delta x < a$; the resulting overlap (and consequent unwrapping of at least one of the two nucleosomes) is characterized by an energetic penalty we refer to as $v(\Delta x)$.

With the above context, the supplementary material of Ref. [36] describes the interaction penalty between two dyads as follows:

> " The available nucleosome positioning data cannot distinguish between different internal nucleosome states. In our model, we therefore consider the internal degrees of freedom as equilibrated and calculate the effective nucleosome-nucleosome interaction. To that end, we consider two neighboring nucleosome particles with a given distance Δx between their dyad positions and sum over the Boltzmann weights of all states compatible with this distance, defining the partition function
>
> $$Z(\Delta x) = \sum_{r,l=0}^{w} \Theta(\Delta x - 2w + r + l) e^{-(r+l)\epsilon} \,, \tag{2.4}$$
>
> where we use the convention $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ otherwise for the Heaviside function. The effective interaction free energy $v(\Delta x)$ between two neighboring particles is then given by
>
> $$\frac{v(\Delta x)}{k_{\mathrm{B}}T} = -\log \frac{Z(\Delta x)}{Z(2w+1)} \,, \tag{2.5}$$
>
> where the normalization factor $Z(2w+1)$ ensures that $v(\Delta x \geq 2w+1) = 0$. When the unwrapping penalty is significant, i.e. in the relevant limit $[\epsilon w \gg 1]$, we can obtain a simple approximation form for the interaction $v(\Delta x)$ by extending the sum in Eq. 2.6 to infinity,
>
> $$\begin{aligned} Z(\Delta x) &\approx \sum_{r,l=0}^{\infty} \Theta(\Delta x - 2w + r + l)\, e^{-(r+l)\epsilon} = \sum_{n=0}^{\infty} \sum_{j=0}^{n} \Theta(\Delta x - 2w + n)\, e^{-n\epsilon} \\ &= \sum_{n=2w-\Delta x+1}^{\infty} (n+1)\, e^{-n\epsilon} = \left(1 - \frac{\partial}{\partial \epsilon}\right) \sum_{n=2w-\Delta x+1}^{\infty} e^{-n\epsilon} \\ &= \left(1 - \frac{\partial}{\partial \epsilon}\right) \frac{e^{-(2w-\Delta x+1)\epsilon}}{1 - e^{-\epsilon}} \,. \end{aligned} \tag{2.6}$$
>
> Calculating the derivative and then substituting into Eq. 2.5 leads to the desired approximation for the interaction free energy,
>
> $$\frac{v(\Delta x)}{k_{\mathrm{B}}T} \approx (a - \Delta x)\epsilon - \log\left[1 + (a - \Delta x)\left(1 - \mathrm{e}^{-\epsilon}\right)\right] \quad \text{for} \quad \Delta x \leq a = 2w+1 \,, \tag{2.7}$$

valid for nucleosome distances Δx not too much smaller than the interaction range $a = 2w + 1$. "([36], Supporting Information).

The values of $v(x)$ can then be used to supplement the matrix from Eq. 2.1 to yield the 'soft-core nucleosome gas' (SoNG) transfer matrix:

$$T_{\mathrm{SoNG}} = \begin{pmatrix} 1 & \mathrm{e}^{\mu} & 0 & 0 & \dots & 0 \\ 0 & \mathrm{e}^{\mu - v(1)} & 1 & 0 & & 0 \\ 0 & \mathrm{e}^{\mu - v(2)} & 0 & 1 & & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & \mathrm{e}^{\mu - v(2w)} & 0 & 0 & & 1 \\ 1 & \mathrm{e}^{\mu} & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{2.8}$$

Where the second column describes the state of a given lattice site, given that the neighboring site (to the right) harbors the beginning of a nucleosome; Δx then depends on the state at the present position. Note that in the limit $v(x) \to \infty \, \forall \, x < a$, Eq. 2.8 reduces to Eq. 2.1. As in the hard-exclusive transfer matrix shown in Eq. 2.1, Eq. 2.8 is not unique. For example, a useful alternative formulation is

$$\tilde{T}_{\mathrm{NN}} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & & 0 & 0 & 0 & & 0 \\ \vdots & & & \ddots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 & 0 & 0 & & 0 \\ 0 & 0 & e^{\mu_N - v(w)} & e^{\mu_N - v(w-1)} & \dots & e^{\mu_N - v(1)} & e^{\mu_N} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 1 & & 0 \\ \vdots & & & & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & & & 0 & 0 & & 1 \\ 1 & 1 & e^{-v(2w)} & e^{-v(2w-1)} & \dots & e^{-v(w+1)} & 0 & 0 & \dots & 0 \end{pmatrix} \cdot \begin{matrix} 0 \\ 1 \\ \vdots \\ w \\ w+1 \\ w+2 \\ \vdots \\ 2w-1 \\ 2w \end{matrix} \tag{2.9}$$

where the state numbers of the matrix are indicated sequentially to the right of the matrix for clarity. Although Eq. 2.8 is more transparent at first glance, Eq. 2.9 will form the basis for incorporating other binding agents into the system in Chapter 3 (the subscript 'NN' signifies exclusively nucleosome-nucleosome interactions, the only binding factor considered in this chapter.)

As with the hard-exclusive case, the transfer matrix defines the partition sum of a lattice array (in this case, a segment of DNA) with periodic boundary conditions as $Z = \mathrm{Tr}\left\{T^L\right\}$. Explicit calculation of configuration probabilities can now be carried out in a more general way.

2.3 Configuration Probabilities

The above discussions have described how to explicitly calculate the partition sum of various 1-D systems using the transfer matrix, but have merely alluded to how the probabilities

of various states can be explicitly determined. Treating nucleosomes as 'located' at their left-most end, we introduce the projection matrix Fs with elements,

$$Fs_{ij} = \delta_{i1}\delta_{ij}, \tag{2.10}$$

where s denotes the fact that Fs is composed in the state basis illustrated in Fig. 2.3, or Fig. 2.2. In either the HaNG or SoNG case, the number of lattice sites in state 1 is then the number of nucleosomes in the array.

From this point, for notational convenience, a general transfer matrix T is used; the following derivation applies equally well to T_{HaNG} or T_{SoNG}[1]. With the above in mind, the probability of a periodic array of length L containing a nucleosome at a given position is then $P_1 = \frac{\mathrm{Tr}\{Fs\cdot T^L\}}{\mathrm{Tr}\{T^L\}}$. Likewise, the two-particle correlation function –i.e. the probability that a particle is found at position x given that there is one also at position 0– can be written as

$$\rho(x|0) = \frac{\mathrm{Tr}\left\{F_s T^x F_s T^{L-x}\right\}}{\mathrm{Tr}\left\{F_s T^L\right\}}. \tag{2.11}$$

Although Eq. 2.11 exactly defines the two-particle correlation function in principle, applying this result to the kind of large systems that are of interest to us in studying long stretches of genomic data requires further analytic work. For that reason, we note that the trace of a matrix is invariant under basis transformations and define P as the matrix of column-wise eigenvectors of T. Naturally, $P \cdot P^{-1} = I$, and $P^{-1} \cdot T \cdot P = D$, where D is the diagonal matrix of T's eigenvalues in decreasing order -i.e. $|\lambda_i| > |\lambda_{i+1}|\ \forall i$. The transformation of Fs into the eigenbasis Fe is, likewise, $Fe = P^{-1} \cdot Fs \cdot P$. The denominator of Eq. 2.11 can then be simplified as follows:

$$\begin{aligned}
\mathrm{Tr}\left\{Fs \cdot T^L\right\} &= \mathrm{Tr}\left\{P \cdot Fe \cdot P^{-1} \cdot \left(P \cdot D \cdot P^{-1}\right)^L\right\} \\
&= \mathrm{Tr}\left\{P \cdot Fe \cdot (D)^L \cdot P^{-1}\right\} \\
&= \mathrm{Tr}\left\{Fe \cdot D^L\right\} \\
&= \sum_{\nu=0} \left(Fe_{\nu\nu}\lambda_\nu^L\right) \qquad (2.12)
\end{aligned}$$

$$(2.13)$$

using the fact that $\mathrm{Tr}(B^{-1}AB) = \mathrm{Tr}(A)$. Likewise, the numerator in equation 2.11 becomes:

$$\begin{aligned}
\mathrm{Tr}\left\{F_s T^x F_s T^{L-x}\right\} &= \mathrm{Tr}\left\{P \cdot Fe \cdot P^{-1} \cdot \left(P \cdot D \cdot P^{-1}\right)^x \cdot P \cdot Fe \cdot P^{-1} \cdot \left(P \cdot D \cdot P^{-1}\right)^{L-x}\right\} \\
&= \mathrm{Tr}\left\{P \cdot Fe \cdot (D)^x \cdot Fe \cdot (D)^{L-x} \cdot P^{-1}\right\} \\
&= \mathrm{Tr}\left\{Fe \cdot D^x \cdot Fe \cdot D^{L-x}\right\}
\end{aligned} \tag{2.14}$$

[1]The calculation here applies also for $\tilde{T}_{\mathrm{NN}}$ from Eq. 2.9, provided nucleosomes are 'located' at their dyad, and the projection matrix is redefined as $Fs_{ij} = \delta_{i,w+1}\delta_{ij}$.

From this point, element-by-element inspection is presented in appendix A.1 to yield

$$\rho(x|0) = \frac{1}{Fe_{00}} \sum_{\nu=0} Fe_{0\nu} Fe_{\nu 0} \left(\frac{\lambda_\nu}{\lambda_0}\right)^x. \tag{2.15}$$

Taking the limit $x \to \infty$, Eq. 2.15 yields the average density $\overline{\rho}$

$$\lim_{x\to\infty} \rho(x|0) = \overline{\rho} = Fe_{00}. \tag{2.16}$$

Eq.s 2.15 and 2.16 are valid for nucleosome calculations with arbitrary neighbor interaction, including the limit of the SoNG model $v(\Delta x) \to \infty \forall x < a$, which is identical to the HaNG model. In this latter case, Eq. 2.8 reduces to Eq. 2.1, and the Tonks gas case emerges.

Eq. 2.15 suffices as a prediction of nucleosome density in the presence of a nucleosome that is strongly-positioned, such as the +1 nucleosome downstream from the TSS. However, the values of Fe and λ_ν depend on μ, a, and in the SoNG case, ε. Constraining these values is accomplished via least-squares residual from experiment as described in the following section.

2.4 Best-Fit Optimization

We are now in a position to apply the theoretical models above to experimental data, although doing so requires constraining the fit parameters[2] $a = 2w + 1$, μ, and in the SoNG case, ε. This is done by minimizing least-squared deviation from experimental data. As with any such simulation, a minimum of fit parameters is desirable in order to choose the most parsimonious model, and the resulting values should not be outside the range of physical plausibility.

At first glance, it may seem as though the SoNG model is in a weaker position by this criteria, owing to the additional fit parameter ε. However, the opposite turns out to be the case when data from a dozen different species are considered simultaneously, as in Ref.[36]. Here, it is shown that reproducing experimental data with the HaNG model requires unique values of both μ and a for each species of yeast –one example of these fits is shown for *C. albicans* in Fig. 2.4A. While individually fitting μ for each species is justified on the basis of different physiological densities (for both the HaNG and SoNG models), fitting the footprint a to each species individually is somewhat more dubious considering the uniform physical characteristics of nucleosomes across species. Moreover, a scatter plot of the HaNG best-fit values of a and $\bar{\rho}$ (which can be mapped one-to-one with μ for any given a) shows a correlation between the two, as seen in Fig. 2.4(B), which implies that the *effective* size of the nucleosome footprint is density dependent, consistent with the SoNG model.

The SoNG model, on the other hand, is much more general. In this case, a single value of a and ε can be fit to all species simultaneously, and is thus referred to as 'unified' in Fig. 2.4C. Despite this reduced freedom in fit parameters (N+2, in total, for N distinct species compared to $2N$ for the HaNG model), the SoNG model still exhibits lower least-squared deviation from experiment (see Ref. [36], supporting information). The lone exception to this success is the

[2] the normalization parameter α mentioned in Section 1.1 is assigned analytically, see Ref. [36], supporting information

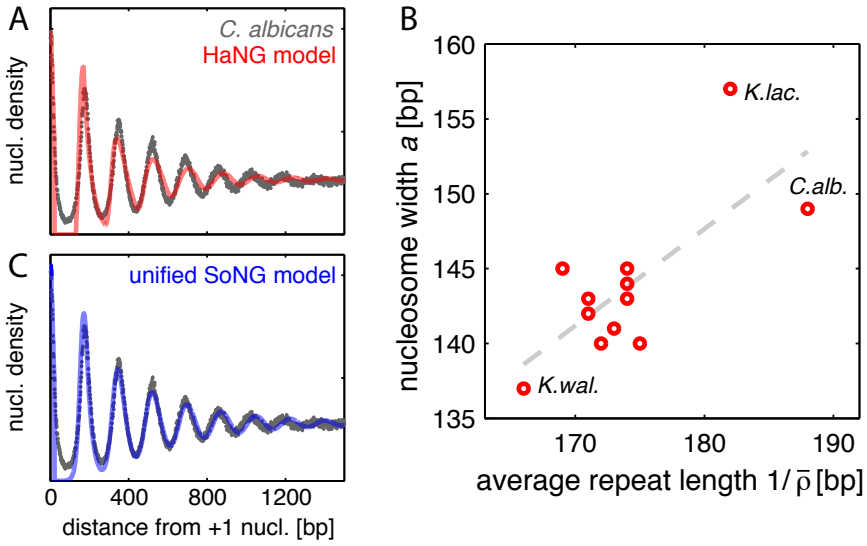


Figure 2.4: Equilibrium nucleosome density downstream from a strongly positioned +1 barrier for (a) hard and (b) soft-interacting nucleosomes. Each model is fit to a set of parameters, however hard nucleosomes exhibit a correlation in the best-fit parameter results that seem to indicate a correlation between density and effective size. This observation would seem to imply that the effective size has some degree of plasticity. Adapted from Fig. 2 of Ref. [36].

species *K. lactis*, which exhibits an anomalously high nucleosome repeat length, for reasons that remain unclear.

Taking best fit parameters with experimental data, the SoNG footprint size $a = 165$ is consistent with the canonical 147 bp binding length with ≈20bp of linker DNA, and happens to also coincide very closely with early predictions of the nucleosome size[10]. The same fitting procedure also produces an optimal binding energy of $\varepsilon = 0.15 k_B T$/bp, consistent with independent experimental measurement[44].

Our derivation of the interaction potential $v(\Delta x)$ assumed competition between histones for DNA binding at the bp level, while other theoretical studies [49] have explicitly included a tendency toward 10-bp unwrapping based on the observation that linker lengths are preferentially quantized[50]. Here, it is important to note that ε is intended only to represent the *average* unwrapping penalty of nucleosomes per bp without rigorously exploring the state-space of possible bonds between helical DNA and the adjacent histone surface. In addition to this, two further subtleties of this model and corresponding fitting procedure have been omitted from the above discussion.

The first relates to the precision of +1 nucleosome position used as a 'barrier'. Despite the relatively strongly positioning of the +1 nucleosome, its finite positioning uncertainty necessitates taking a finite 'spread' of Eq. 2.15 to define a reference point for the barrier at $x = 0$. Details such as this (including, e.g. the domain of analysis downstream from each genes, etc.) are discussed in more detail in the supporting information of Ref. [36], but are not essential to understand the main findings.

Secondly, the above models implicitly assume that the oscillatory pattern in nucleosome

density near the +1 barrier is the result of a gas-like 'pressure' that stacks nucleosomes into regular arrays as they are preferentially adsorbed. This description fails to account for the preservation of the pattern under conditions of reduced-density histone measurements that has been shown experimentally [18]. Separate calculations accounting for these observations form the basis of the next section.

2.5 Kinetic Simulation

While the predictions of the Transfer matrix calculation shown above are excellent predictors of equilibrium nucleosome distribution near a boundary at physiological densities, they are explicitly equilibrium calculations done in a closed, energy conserving system.

Biological cells, however, are much more dynamic, and continually make use of ATP-consuming remodelling machines to regulate the arrangement of histones on the DNA[25], resulting in a *steady state*, but not necessarily equilibrium pattern. The previous section supposes that the strongly-oscillating density pattern of nucleosomes near the beginning of genes, as shown in Fig. 2.4 is the result of an effective 'pressure' from particles being crammed onto the one-dimensional substrate. It would stand to reason, then, that a reduction in the overall density of particles would alleviate this pressure and remove the pattern. In fact, experimental studies have shown that even with a substantial reduction in histone density, density oscillations downstream from genetic promoter regions persist[18].

To account for this observation, a Monte-Carlo-based kinetic simulation is employed based on the Gillespie algorithm, a resource which has become part of the standard set of tools in statistical physics [51]. In this simulation, a lattice of length L is used to represent DNA with all sites available for nucleosome formation with histones from bulk. The resulting event-based dynamics use rates of addition, removal and induced sliding (passive sliding rates are negligible) determined by the Boltzmann factors of the energetic difference between the two states. For example, Fig. 2.5 illustrates states 1 and 2 with internal energy V_1 and V_2 respectively; these latter quantities include all chemical and interaction energies inherent in the corresponding state. The transition rates between the two, $r_{1\to2}$, and $r_{2\to1}$ satisfy detailed balance:

$$\frac{r_{1\to2}}{r_{2\to1}} = e^{\frac{V_1-V_2}{k_BT}}. \tag{2.17}$$

Provided all sites in the array allow for reversible adsorption with rates analogous to Eq. 2.17, and provided a single 'barrier' particle is fixed in position at $x = 0$, the resulting time-average of particle occupation throughout the system is identical to the equilibrium results determined by the transfer matrix in Eq. 2.15. This method, however, allows for the addition of non-equilibrium processes. For example, remodellers that expend energy to serve a spacing role between adjacent nucleosomes by direct interaction between two neighboring nucleosomes have been proposed experimentally[52]. This is just one of many possible models, and there are also other remodeller mechanisms that promote regular spacing[53]. Based on this mechanism, the above kinetic simulation incorporates an additional rate of remodeling k_{rm} which serves to reproduce the steady-state oscillatory density pattern (and long-range heterogeneity) of nucleosomes at low average densities that is seen experimentally[18]. However, it

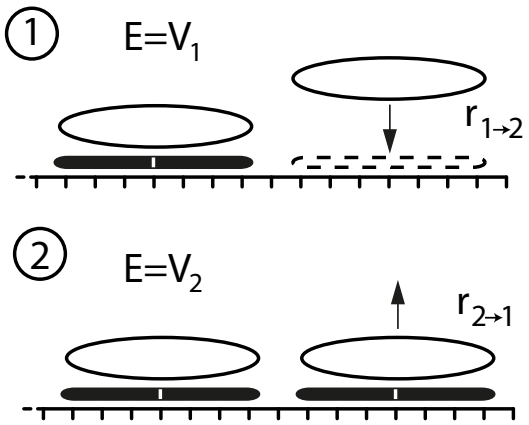


Figure 2.5: Example of two states of an array with transitions representing addition or removal of the particle to the right.

produces negligible change to the density pattern at physiological densities at the same level of activity[36]. Fig. 2.6 illustrates the relative change in density patterns introduced by such remodellers under conditions of different average densities

Taken together, the above results summarize the major findings from reference [36] concerning the nucleosome density pattern downstream from a barrier for 12 species of yeast using the HaNG and SoNG models; the full report contains a more comprehensive description that concludes this chapter.

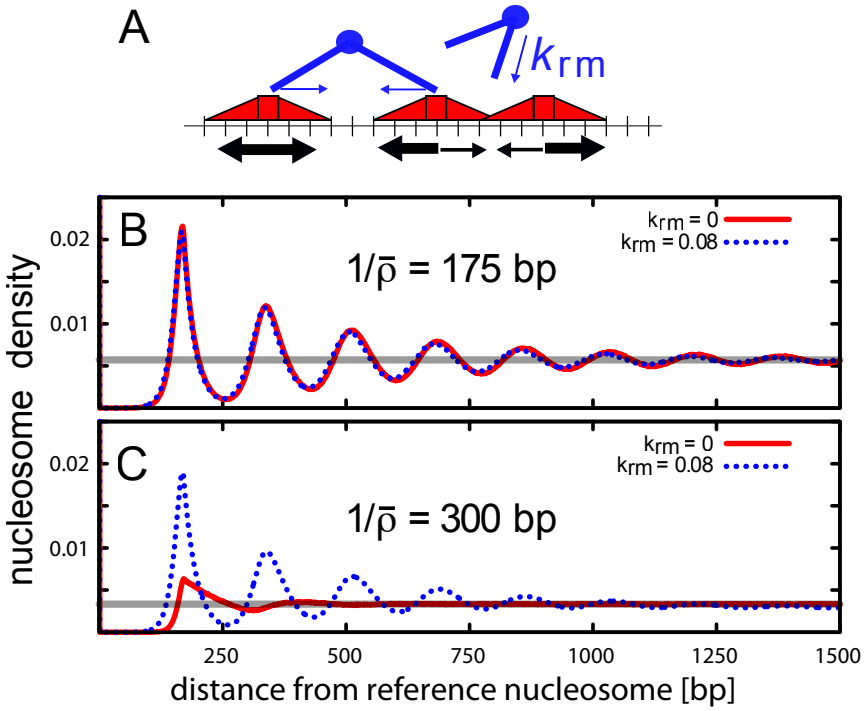


Figure 2.6: Adapted from Ref.[36]. Time averaged Gillespie simulations of 2-nucleosome correlation function at high(b) and low(c) average nucleosome density with and without remodelling based on an experimentally proposed remodeling mechanism[52] sketched in (a). At high density, the remodellers have little effect, while at low density the strongly positioned +1 nucleosome creates a kind of 'seed' for clustering that preserves patterning.

Toward a unified physical model of nucleosome patterns flanking transcription start sites

Wolfram Möbius[a], Brendan Osberg[b], Alexander M. Tsankov[c], Oliver J. Rando[d], and Ulrich Gerland[b,1]

[a]Department of Physics, Harvard University, Cambridge, MA 02138; [b]Arnold Sommerfeld Center for Theoretical Physics and Center for Nanoscience, Department of Physics, Ludwig-Maximilians-Universität, 80333 Munich, Germany; [c]Broad Institute of MIT and Harvard, Cambridge, MA 02142; and [d]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605

Edited by Steven Henikoff, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved February 14, 2013 (received for review August 15, 2012)

Recent genome-wide maps of nucleosome positions in different eukaryotes revealed patterns around transcription start sites featuring a nucleosome-free region flanked by a periodic modulation of the nucleosome density. For *Saccharomyces cerevisiae*, the average in vivo pattern was previously shown to be quantitatively described by a "nucleosome gas" model based on the statistical positioning mechanism. However, this simple physical description is challenged by the fact that the pattern differs quantitatively between species and by recent experiments that appear incompatible with statistical positioning, indicating important roles for chromatin remodelers. We undertake a data-driven search for a unified physical model to describe the nucleosome patterns of 12 yeast species and also consider an extension of the model to capture remodeling effects. We are led to a nucleosome gas that takes into account nucleosome breathing, i.e., transient unwrapping of nucleosomal DNA segments. This known biophysical property of nucleosomes rationalizes a "pressure"-induced dependence of the effective nucleosome size that is suggested by the data. By fitting this model to the data, we find an average energy cost for DNA unwrapping consistent with previous biophysical experiments. Although the available data are not sufficient to reconstruct chromatin remodeling mechanisms, a minimal model extension by one mechanism yields an "active nucleosome gas" that can rationalize the behavior of systems with reduced histone–DNA ratio and remodeler knockouts. We therefore establish a basis for a physical description of nucleosome patterns that can serve as a null model for sequence-specific effects at individual genes and in models of transcription regulation.

nucleosome maps | chromatin structure | quantitative biology

Chromatin is a highly dynamic object and the substrate for molecular processes such as transcription, gene regulation, and DNA replication. Finding adequate model representations for chromatin is challenging, because static structural models are insufficient, and the data required to construct molecular-scale chromatin movies are not within reach. However, conformation-capturing techniques combining biochemical cross-linking and whole-genome mapping provide intriguing statistical information about chromatin conformations. Proximity-based cross-linking of different DNA loci probes the large-scale conformations of chromosomes (1), whereas histone–DNA cross-linking probes conformations on the scale of the 10-nm chromatin fiber (2–5). Such experiments collect a large number of molecular interactions from a population of cells, thereby taking samples from the underlying conformational probability distributions. Hence, these methods inherently lead to a statistical representation of chromatin, which does not explicitly describe the dynamics but instead describes the resulting conformational distribution. A key challenge then is to identify biophysical models that are consistent with the observed conformational distribution.

Here, we consider this question for the statistical distribution of nucleosomes along the DNA. Genome-wide nucleosome-mapping experiments yield a robust pattern upon averaging over a large set of genes aligned at the transcription start site (2–5). It displays a "nucleosome-free region" (NFR), roughly 200 bp wide, and an oscillatory flanking pattern in the direction of transcription. A pattern of this qualitative form has been identified in several multicellular organisms, but is best studied in yeasts (6). The NFRs are likely functionally important, because they affect the access to DNA target sites for factors that regulate and initiate transcription (3–5). Our focus here is on the flanking pattern, a periodic modulation of the nucleosome density, with each peak corresponding to a single nucleosome and an amplitude that decays with distance from the NFR. The average over genes removes gene-specific features and exposes generic physical properties of the 10-nm fiber (7, 8). For the yeast *Saccharomyces cerevisiae*, it was shown (8) that realignment of the genes by their +1 nucleosome (first nucleosome downstream of the NFR) yields a pattern that is quantitatively compatible with that of the barrier nucleosome model (9) based on the "statistical positioning" mechanism (10). Moreover, realigning the genes by their −1 nucleosome (first nucleosome upstream of the NFR) also revealed an upstream oscillatory pattern, which is compatible with that of the same quantitative model.

However, this simple physical interpretation of the pattern is challenged by two sets of experimental observations: First, the quantitative characteristics of the in vivo pattern vary considerably from species to species, even within a set of yeast species (6), raising the question of whether this variation can indeed be understood in simple physical terms. The general physical framework [within which the Kornberg–Stryer model (10) constitutes a special case] is that of one-dimensional (1D) gas systems, consisting of interacting particles confined to a line (11, 12). Within this framework, is there a single model that consistently explains all patterns? Second, recent experiments have shown in vitro (13) and in vivo (14) that the formation of the native nucleosome pattern requires the action of remodeling enzymes. Moreover, reduction of the histone–DNA ratio in vitro (13) or in vivo (15) does not lead to a concomitant increase in the typical nucleosome spacing, as would be expected for statistical positioning. Do these observations invalidate the entire 1D gas framework or just the specific Kornberg–Stryer model?

Here, we address these questions, using a data-driven approach, based on whole-genome nucleosome maps for 12 yeast species (6) (Fig. 1). We find that the species-to-species variation of the patterns cannot be naturally explained within the Kornberg–Stryer model. In particular, it cannot rationalize a conspicuous trend in the data: The effective width of the nucleosome core particle decreases as the mean nucleosome density increases. This observation leads us to an extension of the model that takes into account nucleosome breathing (16–20), the spontaneous transient

Author contributions: W.M., B.O., O.J.R., and U.G. designed research; W.M., B.O., and U.G. performed research; W.M., B.O., and A.M.T. analyzed data; and W.M., B.O., O.J.R., and U.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: gerland@lmu.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1214048110/-/DCSupplemental.

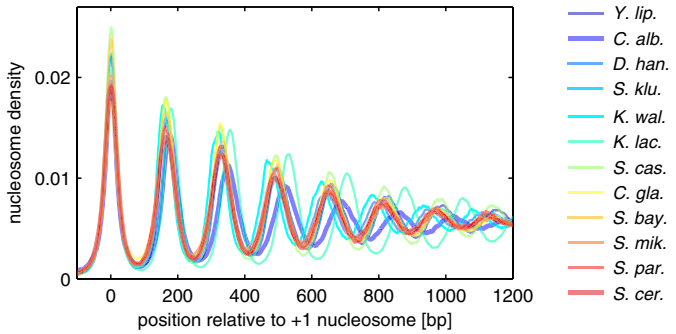


Fig. 1. Average nucleosome density patterns downstream from the +1 nucleosome for 12 yeast species, based on the data of ref. 6. To estimate the physical nucleosome density and enable direct comparison across species, the raw data are normalized by an estimated sequencing depth; see *SI Appendix* for details. To facilitate comparison, the patterns are smoothed by a running average over 11 bp.

unwrapping of nucleosomal DNA from either end of the fully wrapped nucleosome "ground state" represented by the crystal structure (21). This "soft-core" nucleosome gas model has a crucial new biophysical parameter, the mean energy ϵ per base pair required to unwrap nucleosomal DNA. It rationalizes the density dependence of the effective nucleosome width as pressure-induced transient unwrapping of nucleosomes. This explanation does not require rapid equilibration of nucleosome positions within single cells, because the data provide snapshots of independent configurations from many cells. Interestingly, the value for the unwrapping energy ϵ that we infer from the in vivo nucleosome maps agrees with the range of values obtained from biophysical experiments with single-nucleosome core particles. We also show that the soft-core nucleosome gas model can be reconciled with the recent experiments (13–15) probing the dependence on the histone–DNA ratio and the role of remodelers: Adding a dinucleosome-remodeling mechanism to the model does not significantly affect the native in vivo pattern, but renders the typical nucleosome spacing insensitive to the histone–DNA ratio.

Model

Our analysis is based on the working hypothesis that nucleosome arrangements in the 10-nm chromatin fiber can be appropriately described, on a coarse-grained physical level, within the theory of 1D interacting gas systems (12). This description projects the complex molecular interactions among DNA and histones in 3D onto a 1D configurational space. Here, we assume a free energy function of the form

$$E(\{x_j\}) = \sum_i u(x_i) + \sum_i v(x_{i+1} - x_i), \qquad [1]$$

which assigns an energy value to a collection of nucleosome positions $\{x_j\}$. Each x_j is the position of a nucleosome dyad along the DNA and nucleosomes are indexed by order on the DNA ($x_1 < x_2 < \ldots < x_N$). The total number of nucleosomes, N, is not fixed, but may vary from cell to cell and in time. The potential $u(x)$, which acts on each nucleosome individually, is a genomic free energy landscape subsuming *cis* and *trans* effects biasing nucleosome positions (3). The pair interaction $v(x)$ describes the steric exclusion of nucleosomes as well as other mutual interactions. In Eq. **1**, we have restricted $v(x)$ to neighboring nucleosomes. (Additional longer-range interactions due to higher-order chromatin structure are likely, but their effect on the nucleosome patterns appears to be minor, because we find below that a nearest-neighbor–based interaction already describes the available data well.) The Kornberg–Stryer model, which corresponds to the Tonks gas in physics (11), is obtained when $v(x)$ describes a perfect steric repulsion, i.e., $v(x) = \infty$ for $x < b$ and $v(x) = 0$ otherwise (with $b \sim 147$ bp), whereas the energy landscape is flat, $u(x) = 0$, except at isolated "barrier" positions on the genome where $u(x)$ takes on large positive values.

Although the standard theory of 1D interacting gas systems deals with equilibrium systems (12), it is clear that due to the action of various active remodelers, chromatin is an active system. In general, a single energy function of the type Eq. **1** will therefore not suffice to correctly describe the nucleosome distribution at different histone-DNA ratios. Instead, active mechanisms, each characterized by a 'move' within the space of nucleosome configurations and an associated kinetic rate, then need to be added on top of the equilibrium model (22). For active and equilibrium systems alike, the experimentally relevant observable is the local nucleosome density

$$\rho(x) = \left\langle \sum_i \delta(x - x_i) \right\rangle, \qquad [2]$$

i.e., the probability of finding a nucleosome positioned at x. For the experimental data, the angular brackets $\langle \ldots \rangle$ denote an average over different cells, whereas the average is over the appropriate statistical ensemble in the theoretical model (the grand-canonical ensemble for equilibrium systems and the ensemble of steady-state configurations for active systems).

An important general property of dense 1D interacting gas systems is that local perturbations can produce longer-range effects. Specifically, a local peak in the potential landscape $u(x)$, repulsive or attractive, can trigger an oscillation in $\rho(x)$ over a range of many nucleosomes, due to a statistical "domino effect," where the positioning of a nucleosome constrains its neighbor and the latter its neighbor in turn. In physical terms, oscillatory behavior of $\rho(x)$ results from the interplay of entropy and steric repulsion: Because the same DNA position cannot simultaneously be incorporated into two nucleosomes, the nucleosome footprints on the DNA exclude each other, whereas a large gap between neighboring nucleosomes reduces the configurational space for the remaining nucleosomes and is therefore entropically unfavorable. For *S. cerevisiae*, a parsimonious explanation of the experimental $\rho(x)$ in the vicinity of transcription start sites (TSS) assumes an asymmetrically shaped potential $u(x)$, where the +1 nucleosome is directly positioned with a localized attractive peak (negative potential), whereas the −1 and upstream nucleosomes are statistically positioned by a repulsive potential within the NFR (8). Because the experimental pattern results from an alignment and average over a large set of genes, this shape of $u(x)$ appears to be common to most genes. Additionally, there may well be other features in the genomic $u(x)$ landscape, e.g., also within the coding regions, which vary from gene to gene such that they do not significantly affect the average pattern.

The precise form of the oscillatory pattern depends on the shape of the peak in $u(x)$ that defines the boundary condition for the pattern (8). One can infer the most likely local shape of $u(x)$ that positions the +1 nucleosome from the experimental shape of the +1 peak in the average density. [For simplicity, in our calculations reported below, we take this boundary effect into account via a convolution of the idealized theoretical pattern for a sharply peaked $u(x)$ with the shape of the experimental +1 peak (*SI Appendix*).] In the following, we use the term hard-core nucleosome gas (HNG) model to refer to the Kornberg–Stryer–Tonks model supplemented by this boundary condition ("HNG" is chosen in contrast to the soft-core model introduced in *Soft-Core Nucleosome Gas Model* below). By the logic of Occam's razor, we seek to identify the simplest and most plausible combination of an energy function, Eq. **1**, and active

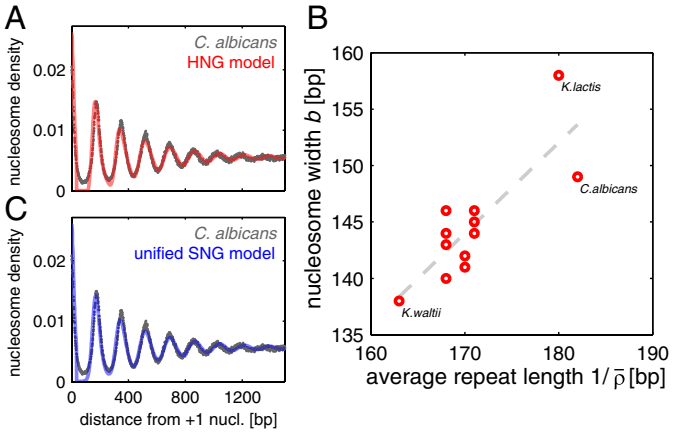


Fig. 2. (*A*) Overlay of the *C. albicans* pattern (gray dots) with its best HNG model fit, where particle size b and mean density $\overline{\rho}$ are independently varied. (*B*) Parameter pairs $(\overline{\rho}, b)$ of the best HNG model for each species (circles). The nucleosome width b displays a positive correlation with the average repeat length $1/\overline{\rho}$ (Pearson's correlation coefficient $r = 0.82$ for all species and 0.78 for all but *K. lactis*). The dashed line is a linear fit to all data. (*C*) Similar to *A*, but with the unified SNG model, where only $\overline{\rho}$ is independently fitted for each species.

mechanism(s) that can rationalize the nucleosome patterns from different species and in vitro experiments.

Results

Quantitative Data Analysis. Fig. 1 shows the average nucleosome pattern flanking the NFR on the downstream side, for each of the 12 different yeast species of ref. 6. Here, instead of aligning at the TSS, we used the most likely +1 nucleosome position at each gene as an alignment point (*SI Appendix*). In terms of the physical model, the +1 nucleosome is our reference particle, whereas the TSS cannot be mapped to a physical feature (8). All of the patterns in Fig. 1 have the same qualitative shape, a decaying oscillation, but the quantitative characteristics differ significantly (6). To test whether these differences are due to variation of gene expression level or gene length, we used the data of ref. 6 to compare, for each species, the patterns of genes with high and low expression levels and the pattern of the long genes with the average over all genes (*SI Appendix*, Figs. S1*B*–S12*B*). This reveals two trends: (*i*) Genes with high expression levels display a pattern that appears "compressed"; i.e., nucleosomes have a smaller spacing than for low expression levels. (*ii*) Longer genes display a more pronounced pattern, i.e., larger peak amplitudes at equal spacing. Although these trends are consistent with previous observations (3–5), both effects are relatively small compared with the species-to-species differences.

To test whether the HNG model can rationalize the different patterns, we fitted it using the average nucleosome density $\overline{\rho}$ as an adjustable parameter. Because $\overline{\rho}$ controls the peak-to-peak distances in the pattern and it is well known that the average nucleosome spacing in coding regions is variable between yeast species (6, 23), a species-to-species variation in $\overline{\rho}$ might already explain the observed patterns. We recovered the previous finding that the HNG model describes the *S. cerevisiae* pattern relatively well (8), as quantified by the mean-square deviation, δ^2. However, the fit was poor for some of the other species, especially *Kluyveromyces waltii*, *Kluyveromyces lactis*, and *Debaryomyces hansenii* (*SI Appendix*, section IV and Figs. S1*A* –S12*A* and S13).

To elucidate this failure of the HNG model, we allowed the effective DNA footprint size of nucleosomes to be adjustable as well (b was previously fixed at 147 bp). With the additional parameter, we obtained good fits for all patterns; see Fig. 2*A* for an example and *SI Appendix* for complete results. As shown in Fig. 2*B*, these fits yield a range of parameter values, with average repeat lengths $1/\overline{\rho}$ between 163 and 182 bp and effective DNA footprint lengths b from 138 to 158 bp. Given that the nucleosome structure has a well-defined 146- to 147-bp length of nucleosomal DNA (21) and the histone sequences are highly conserved, the large spread in b indicates that the model does not fully capture the physical behavior of the 10-nm chromatin fiber. Indeed, Fig. 2*B* shows that the two fitted parameters are not independent (correlation coefficient $r = 0.82$), suggesting that the model misses an essential property of the system. Additional analysis (*SI Appendix*) confirms that this correlation is not an experimental or a computational artifact.

One biophysical property neglected by the HNG model is the dynamic nature of the nucleosome structure: It is well known that the ends of nucleosomal DNA transiently unwrap from the histone core driven by thermal fluctuations (16, 18–20), even in the chromatin context (17). This "nucleosome breathing" reduces the average histone-bound DNA length in a nucleosome below that in the crystal structure. Importantly, physical reasoning suggests that the size of this reduction should depend on the nucleosome density: A high density corresponds to a large effective "pressure" of the nucleosome "gas", which induces more unwrapping. This intuitive picture is indeed applicable even when the dynamics of nucleosome sliding along the DNA are slow, because (*i*) the experimental patterns are generated from many "snapshots" of nucleosome configurations in different cells and (*ii*) the picture depends only on the equilibrium statistics, not the dynamics, of the system, with the unwrapping resulting from an energy–entropy trade-off (see below).

Soft-Core Nucleosome Gas Model. To test whether this physical effect can quantitatively explain the data, we devised a "soft-core nucleosome gas" (SNG) model, which accounts for nucleosome breathing. In the SNG model, nucleosomes have two internal degrees of freedom corresponding to the amount of unwrapped DNA on each side of the dyad (Fig. 3*A*). Unwrapping has a free energy cost $\epsilon > 0$. Consequently, two adjacent nucleosomes at a given distance

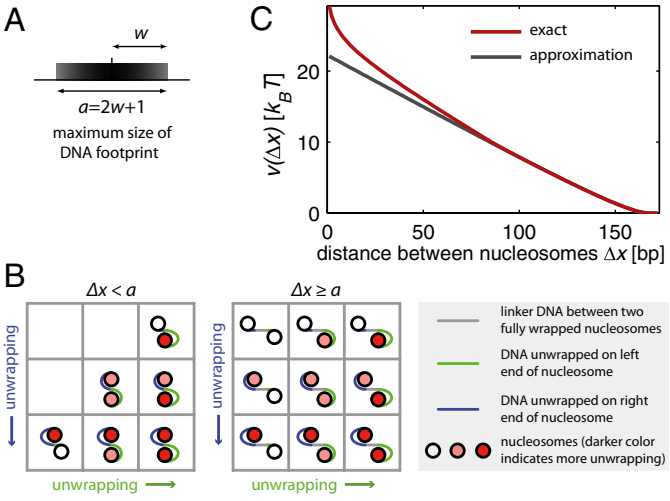


Fig. 3. Soft-core nucleosome gas (SNG) model. (*A*) A single soft-core particle has a maximal footprint a. DNA can unwrap from both ends, up to a maximum of w on either side of the dyad (marked by a vertical line). These "breathing" dynamics lead to a reduced typical footprint size (illustrated by shading). (*B*) Illustration of the different internal states of a nucleosome pair that contribute to the effective nucleosome–nucleosome interaction free energy $v(\Delta x)$ at a given distance Δx between the dyads. (*C*) This interaction is plotted for our consensus parameter values ($w = 83$ bp, $\epsilon = 0.1525 k_B T$/bp). The approximation of Eq. **3** (gray line) matches the exact form (red line) in the relevant regime of small free energy.

$\Delta x = x_{i+1} - x_i$ between the dyads can be in a number of different internal states, as illustrated in Fig. 3*B*. Without any unwrapping, they cannot be closer than a certain minimal distance a. For $\Delta x < a$, at least $a - \Delta x$ bp must unwrap, which can be split between the left and the right particle. A priori, we expect that a is larger than 147 bp, because a configuration without unbound (linker) DNA between two nucleosomes is difficult to achieve given the steric constraints of chromatin. After summing over the internal states, which are not observed in the experiments, we obtain an interaction free energy for neighboring nucleosomes at distance $\Delta x \leq a$,

$$v(\Delta x) \approx (a - \Delta x)\epsilon - k_B T \ln\left[1 + (a - \Delta x)\left(1 - e^{-\epsilon/k_B T}\right)\right], \qquad [3]$$

whereas $v(\Delta x)$ remains 0 for $\Delta x > a$. The first term of this "soft-core potential" is enthalpic, whereas the second term is entropic, accounting for the different internal states. Eq. **3** is an approximate expression valid for $\Delta x \gg 1$ bp (*SI Appendix*). Fig. 3*C* displays both the exact and the approximative $v(\Delta x)$ with representative parameter values. Within the general theory of 1D gas models, our SNG model falls into the class of "Takahashi nearest-neighbor gases" (12). All of its statistical properties can be calculated exactly, using standard techniques from statistical physics. Conceptually similar models were considered by Chou (24) and Teif et al. (25). Lubliner and Segal showed that taking into account interactions between adjacent nucleosomes can improve DNA sequence-based prediction of nucleosome occupancy (26).

We first fitted the SNG model separately to each species, with the average nucleosome density $\overline{\rho}$, the unwrapping energy ϵ, and the interaction range a as adjustable model parameters. The best-fit parameters and the fit quality are indicated with orange bars in Fig. 4, and the individual patterns are shown in *SI Appendix*, Figs. S1*A*–S12*A*. *SI Appendix*, Fig. S13 compares the fit qualities of all of the models considered so far: As expected, the SNG model shows a dramatic improvement over the restricted HNG model (fixed b). It describes the experimental patterns about equally as well as the HNG model with adjustable b. Importantly, the SNG model appears to correctly capture the physical behavior of the system: (*i*) The average repeat lengths $1/\overline{\rho}$ determined from the SNG model agree well with the peak-to-peak distances in the experimental patterns, whereas the repeat lengths determined from the HNG model are significantly longer (*SI Appendix*, Tables S2 and S3). (*ii*) The inferred unwrapping energies ϵ are in excellent agreement with biochemical measurements of DNA unwrapping, which indicate a free energy cost of about 1.5 $k_B T$ for the unwrapping of 10 bp (27). (*iii*) The inferred interaction ranges $a = 2w + 1$ are roughly 20 bp larger than the DNA length of the nucleosome core particle. In other words, already before the DNA footprints of neighboring nucleosomes overlap, they experience a weak repulsive interaction. Physically, such a repulsion arises, e.g., from the geometrical constraints that reduce the conformational space of a nucleosome pair at short distances.

So far, the SNG model appears to share with the HNG model the conceptual problem of species-specific histone properties despite the high degree of histone conservation: Not only the nucleosome spacings, but also the best-fit unwrapping energy and interaction range vary from species to species (Fig. 4). However, intuition suggests that larger particles with strong breathing can yield the same average particle size as smaller particles that breathe little; i.e., a decreased ϵ may be compensated for by an increased w. We therefore asked: Can the histone properties w and ϵ be constrained to be universal for all species without losing the quality of the description? This "unified SNG model" has a substantially reduced number of adjustable parameters for the description of all data compared with the SNG model or the HNG model with variable particle width, because only the average nucleosome density $\overline{\rho}$ remains as a species-specific physical parameter. When performing this constrained and simultaneous optimization, we found that the unified SNG model can describe the data almost equally as well as the SNG model itself, except for the case of *K. lactis*, which we excluded in this analysis (discussed below). We find a unified binding energy of $\epsilon = 0.1525\ k_B T$/bp and a maximum wrapping length of $w = 83$ bp corresponding to a unified interaction range of $a = 167$ bp. The resulting $\overline{\rho}$-values and fit qualities δ^2 are shown in Fig. 4 (blue bars) and the individual patterns as well as the comparison with the HNG models are in *SI Appendix*.

As speculated above, when moving from the SNG to the unified SNG model, an increase in ϵ is typically accompanied by a decrease in a and vice versa without a severe drop in fit quality. This trend is particularly clear in the case of *K. waltii* and *Candida albicans*. In contrast, *K. lactis* is abnormal in this respect, because the individually estimated values for ϵ and a are both exceptionally high, suggesting that nucleosomes in *K. lactis* breathe less or interact over a larger range than in other yeast species. Accordingly, the unified SNG model for the other 11 yeast species describes the *K. lactis* data poorly (Fig. 4 and *SI Appendix*). The abnormal behavior of *K. lactis* could have various mechanistic origins, as discussed below. Because the average nucleosome spacing generally shows a (weak) dependence on gene expression level, one possible

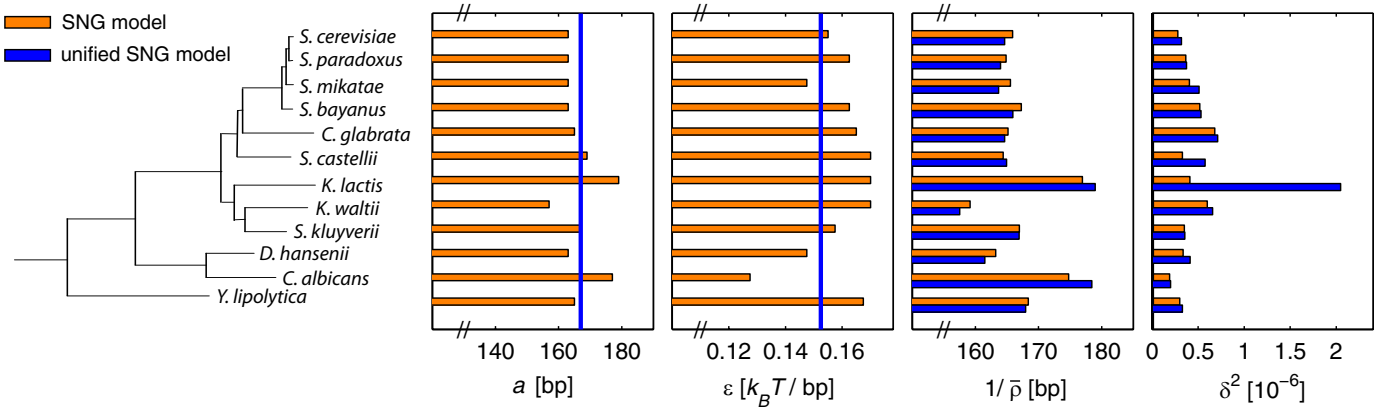


Fig. 4. Bar graph of the parameter values for the species shown in the phylogeny as inferred from the independent (orange) and unified (blue) SNG models. The rightmost column shows the mean-square deviation δ^2 between experimental and model patterns. For the unified SNG model, the interaction range a and binding energy ϵ are global parameters (inferred values indicated by vertical blue lines). The average repeat length $1/\overline{\rho}$ is a species-specific parameter in both cases. *K. lactis* is the only species displaying poor agreement with the unified model and was not included in the optimization of the global parameters.

origin could be related to gene expression. *SI Appendix*, Fig. S14 separately compares the patterns of genes with high and low expression levels between *K. lactis* and *S. cerevisiae*. Because *K. lactis* displays a larger spacing than *S. cerevisiae* at high and low expression levels alike, the abnormal behavior of *K. lactis* is unlikely related to gene expression.

Does the unified SNG model indeed display the pressure-induced unwrapping effect postulated above? To address this question, *SI Appendix*, Fig. S15 shows the effective footprint size a_{eff} (ensemble average over all configurations) as a function of the inverse density $1/\overline{\rho}$ (average repeat length). Over the relevant range of $1/\overline{\rho}$, from 158 to 178 bp for the 11 species included in the unified model, a_{eff} varies between 144 and 151 bp. Hence, we find pressure-induced unwrapping to be a relevant effect. As an aside, we note that spontaneous and pressure-induced unwrapping could also significantly contribute to the recently observed broad distribution of micrococcal nuclease-protected DNA sizes in paired-end sequencing studies (28) (in addition to the existence of other types of particles).

Although the SNG model yields a surprisingly accurate description of the different in vivo nucleosome patterns within the same physical model, it leaves open a number of important questions. An immediate question is why *K. lactis* does not fit into the unified model, especially because it is not at an extremity of the phylogenetic tree (Fig. 4). It was previously found that the nucleosome repeat length of *K. lactis* is 16 bp longer than that of *S. cerevisiae* (23), which was speculated to be due to an abundant use of the linker histone H1 in *K. lactis*. This would indeed explain a concomitant increase of average nucleosome spacing and apparent nucleosome width. Whereas the mRNA level of the H1 ortholog in *K. lactis* is similar to that of H1 in *S. cerevisiae* (6), a significant difference in H1 abundance could arise at the post-transcriptional level. Alternatively, the behavior of *K. lactis* could be rationalized by histone modifications that make DNA unwrapping energetically more costly and the apparent nucleosome width larger than in the other yeast species (Fig. 4).

Remodeling Enzymes. We have seen that under in vivo conditions the nucleosomes effectively behave like a 1D gas of interacting particles, at least with respect to the observables considered so far. This does not imply that the 10-nm chromatin fiber actually is an equilibrium system with such simple interactions. The true mix of molecular mechanisms that produce the effective in vivo behavior is only beginning to be disentangled experimentally (29) and in physical models of nucleosome organization (22). An in vitro reconstitution of in vivo-like nucleosome patterns was shown to be dependent on ATP and on factors acting in *trans* (13). Moreover, a reduction of the histone–DNA ratio by 50% in vitro (13) and ∼30% in vivo (15) surprisingly left the typical nucleosome spacing close to TSSs largely unchanged. This finding is clearly at odds with pure statistical positioning, which predicts an increased spacing for reduced nucleosome counts. Instead, Zhang et al. (13) suggest an active packing mechanism mediated by remodeling enzymes, e.g., a dinucleosome packing mechanism of the type suggested for the imitation switch family remodeler ISW1a (30). We now explore the possibility that such a packing mechanism acts in addition to the mechanisms of the SNG model.

Active SNG model. As a minimal extension of the SNG model, we introduce a remodeler that randomly binds, at rate k_{RM}, to two adjacent nucleosomes and actively pulls one of them closer to the other (by 1 bp). This "active SNG model" is illustrated in Fig. 5*A*; see *SI Appendix* for details. We assume the remodeler has a maximal reach such that it can bind if the distance between the nucleosome dyads is up to $4w$. Although it has no intrinsic directional bias, an apparent packing toward the +1 nucleosome can emerge indirectly: The dinucleosome mechanism mediates an attraction between nucleosomes, creating variable-sized nucleosome clusters anywhere on the DNA. A strongly positioned +1 nucleosome seeds such a cluster and effectively pins its position. To illustrate the model behavior, *SI Appendix*, Fig. S16 shows the time evolution of a typical simulation. The emerging average pattern downstream of a well-positioned (+1) nucleosome is shown in Fig. 5*B* for a normal histone–DNA ratio, whereas Fig. 5*C* shows the pattern obtained at a significantly reduced nucleosome count. To expose the effect of the remodeler, Fig. 5 *B* and *C* show not only the pattern with remodeling (dashed line, $k_{RM} = 0.08$ in units of the intrinsic sliding rate of nucleosomes), but also that for $k_{RM} = 0$ (solid line). Strikingly, the remodeler has virtually no effect on the nucleosome pattern at the normal histone–DNA ratio (whereas it is crucial at the reduced ratio to maintain the phasing). Intuitively, the compressing effect of the remodeler is weak compared with the gas pressure of the SNG model at high density. This behavior is not sensitive to the precise choice of k_{RM}, because the nucleosome gas pressure rises steeply at high densities.

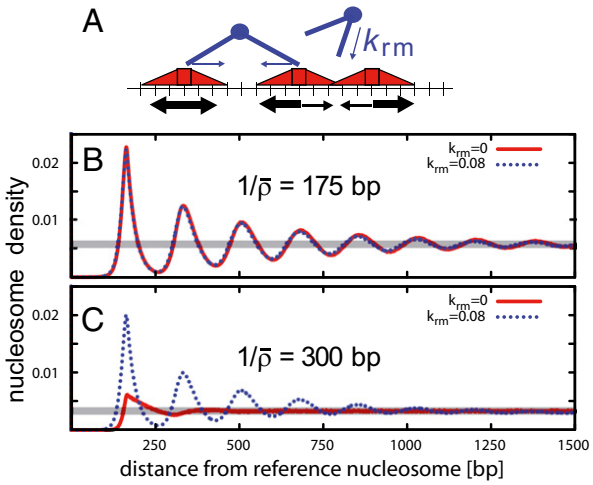


Fig. 5. Active SNG model. (*A*) Illustration of the kinetic model for the coupled dynamics of passive nucleosome sliding and active remodeling. Sliding occurs in 1-bp steps and is constrained by the soft-core nucleosome interaction (relative rates set by the energetics of interaction with neighbors; larger arrows indicate faster processes). Active remodeling is mediated by remodelers acting on dinucleosomes (represented by "tweezer" symbols). A remodeling move brings two neighboring nucleosomes closer to each other by 1 bp, by randomly moving one of the two nucleosomes. (*B* and *C*) At physiological histone density (*B*) remodeling at moderate rates has no effect on the nucleosome pattern, whereas at low density (*C*) the pattern is maintained only with remodeling.

Discussion

The active SNG model is compatible with the preserved nucleosome spacing at reduced histone abundance (13, 15) and the seemingly contradictory finding that the normal in vivo nucleosome pattern is well described by the barrier nucleosome model without explicitly accounting for remodelers (7–9). The model predicts that a reduced histone abundance produces a slight downward slope in the average nucleosome pattern; i.e., the oscillatory density decays to the reduced average value at long distances from the reference nucleosome (*SI Appendix*, Fig. S17). Physically, this is a signature of the formation of nucleosome clusters, which have a statistically distributed cluster size, such that the likelihood of leaving the cluster of the reference nucleosome increases with distance. The in vivo data of Celona et al. indeed suggest this trend; compare figure 7B in ref. 15.

Further clues about the physical mechanisms underlying (yeast) nucleosome patterns are provided by the experimental findings that (*i*) in vitro reconstitution of an in vivo-like nucleosome

pattern requires whole-cell extract and ATP (13) and (*ii*) a combined knockout of several remodelers shows a clear phenotype in the average nucleosome pattern, whereas individual knockouts have mild effects (14). Given that both observations are at a physiological histone–DNA ratio, they can be rationalized kinetically only within the active SNG model: As previously pointed out (22), the intrinsic dynamics of nucleosomes, i.e., fluctuation-induced unbinding and diffusion along DNA, are too slow to equilibrate their positions on the relevant timescales of cell division or in vitro experiments. The remodelers could significantly speed up these dynamics, such that even at the normal histone–DNA ratio an in vivo-like nucleosome pattern cannot be generated in time without active remodelers.

Conclusions

Our results support a physical interpretation of the statistical distribution of nucleosomes around TSSs, where nucleosomes are viewed as particles with a soft repulsive core in a 1D space. Although this description ignores higher-order interactions between nonadjacent nucleosomes, it captures in vivo patterns of different yeast species surprisingly well. The behavior of mutant and in vitro systems as well as the dynamic process of nucleosome pattern formation can be understood by introducing remodeling enzymes into the model, which constitute active components in the 1D nucleosome gas. We proposed one such active component that effectively leads to an attraction between nucleosomes. Interestingly, it can rationalize several key observations in mutant and in vitro systems while retaining the model behavior in the parameter regime corresponding to the in vivo situation. We note that an effective attraction can also be produced through molecular mechanisms other than dinucleosome remodeling. In particular, nucleosome interactions via higher-order chromatin structure could also contribute to this attraction (31–33). Taken together, we believe the active SNG model currently provides the most comprehensive physical description of nucleosome patterns in yeasts, although it is far from a truly mechanistic description of the system. We see several practical uses for the model, e.g., as a physical "null model" for future gene-by-gene analyses of nucleosome positioning and as a component in mechanistic models for transcription regulation in yeast.

ACKNOWLEDGMENTS. We thank Philipp Korber for helpful suggestions. We are indebted to Jonathan Widom for enlightening discussions and greatly saddened by his untimely death. We would like to dedicate this work to him. This work was funded, in part, by the Bavarian Center for Molecular Biosystems, the German Excellence Initiative via the program "Nanosystems Initiative Munich" (NIM), and the Deutsche Forschungsgemeinschaft through SFB 1032. W.M. acknowledges funding by a Leopoldina Postdoctoral Scholarship. B.O. is grateful to the Center for NanoScience and the Elite Network of Bavaria (International Doctorate Program in NanoBioTechnology).

1. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295(5558):1306–1311.
2. Yuan GC, et al. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* 309(5734):626–630.
3. Segal E, Widom J (2009) What controls nucleosome positions? *Trends Genet* 25(8): 335–343.
4. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: Advances through genomics. *Nat Rev Genet* 10(3):161–172.
5. Radman-Livaja M, Rando OJ (2010) Nucleosome positioning: How is it established, and why does it matter? *Dev Biol* 339(2):258–266.
6. Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ (2010) The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol* 8(7):e1000414.
7. Chevereau G, Palmeira L, Thermes C, Arneodo A, Vaillant C (2009) Thermodynamics of intragenic nucleosome ordering. *Phys Rev Lett* 103(18):188103.
8. Möbius W, Gerland U (2010) Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLOS Comput Biol* 6(8):e1000891.
9. Mavrich TN, et al. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 18(7):1073–1083.
10. Kornberg RD, Stryer L (1988) Statistical distributions of nucleosomes: Nonrandom locations by a stochastic mechanism. *Nucleic Acids Res* 16(14A):6677–6690.
11. Tonks L (1936) The complete equation of state of one, two and three-dimensional gases of hard elastic spheres. *Phys Rev* 50:955–963.
12. Mattis DC, ed (1993) *The Many-Body Problem: An Encyclopedia of Exactly Solved Models in One Dimension* (World Scientific, Singapore).
13. Zhang Z, et al. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* 332(6032):977–980.
14. Gkikopoulos T, et al. (2011) A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* 333(6050):1758–1760.
15. Celona B, et al. (2011) Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biol* 9(6):e1001086.
16. Polach KJ, Widom J (1995) Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J Mol Biol* 254(2): 130–149.
17. Poirier MG, Bussiek M, Langowski J, Widom J (2008) Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol* 379(4):772–786.
18. Tims HS, Gurunathan K, Levitus M, Widom J (2011) Dynamics of nucleosome invasion by DNA binding proteins. *J Mol Biol* 411(2):430–448.
19. Möbius W, Neher RA, Gerland U (2006) Kinetic accessibility of buried DNA sites in nucleosomes. *Phys Rev Lett* 97(20):208102.
20. Koopmans WJA, Buning R, Schmidt T, van Noort J (2009) spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophys J* 97(1):195–204.
21. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648):251–260.
22. Padinhateeri R, Marko JF (2011) Nucleosome positioning in a model of active chromatin remodeling enzymes. *Proc Natl Acad Sci USA* 108(19):7799–7803.
23. Heus JJ, Zonneveld BJ, Bloom KS, de Steensma HY, van den Berg JA (1993) The nucleosome repeat length of Kluyveromyces lactis is 16 bp longer than that of Saccharomyces cerevisiae. *Nucleic Acids Res* 21(9):2247–2248.
24. Chou T (2003) An exact theory of histone-DNA adsorption and wrapping. *Europhys Lett* 62:753–759.
25. Teif VB, Ettig R, Rippe K (2010) A lattice model for transcription factor access to nucleosomal DNA. *Biophys J* 99(8):2597–2607.
26. Lubliner S, Segal E (2009) Modeling interactions between adjacent nucleosomes improves genome-wide predictions of nucleosome occupancy. *Bioinformatics* 25(12): i348–i355.
27. Anderson JD, Widom J (2000) Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol* 296(4):979–987.
28. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S (2011) Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci USA* 108(45): 18318–18323.
29. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78:273–304.
30. Yamada K, et al. (2011) Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature* 472(7344):448–453.
31. Blank TA, Becker PB (1995) Electrostatic mechanism of nucleosome spacing. *J Mol Biol* 252(3):305–313.
32. Emanuel M, Radja NH, Henriksson A, Schiessel H (2009) The physics behind the larger scale organization of DNA in eukaryotes. *Phys Biol* 6(2):025008.
33. Chereji RV, Morozov AV (2011) Statistical mechanics of nucleosomes constrained by higher-order chromatin structure. *J Stat Phys* 144(2):379–404.

3 An Equilibrium Two-Component Gas Model

While histones are the most prevalent binding agent along the genome (for yeast as well as other eukaryotes), other proteins, such as transcription factors (TF), are also present and compete for binding sites along the DNA. Transcription factor binding, for example, has been shown to be correlated with the nucleosome free region (NFR) in the promoter regions, not only due to direct competition for binding sites, but also partly due to their recruitment of remodellers [54, 55]. The distribution of DNA-binding proteins in the promoter region has been shown to play an important regulatory role in transcription initiation[28, 56, 57, 20, 58, 59]. Furthermore, it has been demonstrated that changing cell conditions triggers changes in the binding pattern near the promoter regions[60]. The interplay between transcription and chromatin structure in the promoter region has even been proposed as a means of designing synthetic gene circuits[61].

In this chapter we extend the one-dimensional model of competitive site-binding into a two-component lattice gas model to reflect the interaction of smaller, usually specifically binding, TFs. To motivate this work, experimental data of shifted nucleosome profiles was gathered from published data and compared to TF binding positions in the course of Ref. [62]. The competitive binding of TFs and nucleosomes is particularly relevant in the context of our SoNG model from the previous chapter since it is known that transient unwrapping from the end of the nucleosome can allow for TF binding to positions that would otherwise be occluded in the HaNG model[40, 63, 45].

It has also been shown that multiple TFs whose binding sites overlap with the same nucleosome bind cooperatively, as they work together to displace a single nucleosome [64]. Calculations presented below suggest that similar effective cooperativity can even exist between transcription factors whose binding sites are separated by several intervening nucleosomes. As we will see below, shifted nucleosome patterns under the influence of one TF can affect the availability of sites for others. We refer to this interplay as nucleosome-mediated cooperativity between TFs and provide a quantification (first presented in Ref. [65]) of this effect.

3.1 Equilibrium Calculations

To incorporate TFs into a theoretical model of lattice binding, it is necessary first to construct an interaction potential between the two species, just as the nucleosome-nucleosome interaction potential was used in the previous chapter. Previous studies of TF-nucleosome interaction have assumed hard-exclusion in DNA binding (analogous to our HaNG model)[54]. It has been demonstrated experimentally, however, that site exposure (and availability for TF-binding) of nucleosomal DNA due to transient unwrapping decreases incrementally from

the outer edge of the nucleosome inward[45]. Thus, for TFs binding competitively with an adjacent nucleosome, we propose a progressive unwrapping penalty that mirrors the underlying physics of the nucleosome-nucleosome SoNG interaction from the previous chapter. Transient unwrapping allows for exposure of certain sites along the DNA with the energetic penalty of broken chemical bonds. One major difference from the SoNG interaction, however, is that we assume TFs are sufficiently tightly bound as to prevent unwrapping; thus, the neighbor interaction involves summing over the set of unwrapping states from only a single side.[1] The effective interaction potential $u(\Delta x)$ between a TF of size m bp whose nearest side is separated from a nucleosome dyad by Δx bp is determined by the sum of the Boltzmann factors of possible binding states of the nucleosome:

$$u(\Delta x) = \begin{cases} -\ln\left[\frac{\sum_{i=0}^{\Delta x-1} e^{i\varepsilon}}{\sum_{j=0}^{w} e^{j\varepsilon}}\right] & \text{if } \Delta x \leq w \\ 0 & \text{if } \Delta x > w \end{cases} \tag{3.1}$$

where w, again, is the length of DNA on one side of the dyad that can bind to the histone surface. There are strong parallels with with Eq. 2.5 from Chapter 2, and Fig. 3.1 shows that the two potentials are similar in shape, but with u acting at half the range.

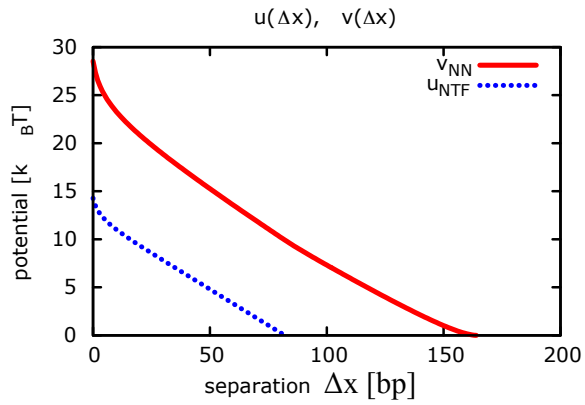


Figure 3.1: Neighbor interaction potential at a distance of Δx bp between two nucleosomes $v_{\mathrm{NN}}(\Delta x)$(red), and a Nucleosome with the nearest edge of a TF $u_{\mathrm{NTF}}(\Delta x)$(blue). To a good approximation, u_{NTF} is similar to v_{NN}, but with half the range.

It is quite straightforward to adapt the Gillespie algorithm of Chapter 2 to include particles of another type, and all of the same principles of reversible adsorption and detailed balance still apply –provided the additional energetic terms described above are included into the global energy change assumed in Eq. 2.17. Notably, however, the TFs are highly specific in their binding behavior, which is reflected in a position-dependent binding energy $\mu_{\mathrm{TF}}(x)$, as opposed to nucleosomes which bind non-specifically with energy μ_{N}. As in Chapter 2, $\mu_{\mathrm{TF}}(x)$ determines the binding rate $r_{+\mathrm{TF}} = e^{\mu_{\mathrm{TF}}(x)-u(\Delta x_l)-u(\Delta x_r)}$ in conjunction with any energetic interactions $u(\Delta x_l)$ and $u(\Delta x_r)$ with the left and right neighbors respectively. For simplicity, we take $\mu_{\mathrm{TF}}(x) \to -\infty$ at non-specific binding positions.

[1] There is, however, no reason to think that the binding energy of nucleosomes per bp ε would differ from the value determined in Chapter 2 via least-squares regression and from experimental results in Ref. [44].

To illustrate the effects of localized occlusion on the nearby nucleosome distribution pattern, we take a system with a fixed nucleosome at position $x = 0$, and measure the effect of adding a TF specific binding site at the downstream position $r = 360$, shown in Fig. 3.2 with $\mu_{\mathrm{TF}}(360) = \mu_{\mathrm{N}} = 9k_BT$. Here, the red data illustrate nucleosome density –in the absence of any TFs– similar to what was seen in the previous chapter, while the blue data show the resulting shift in this pattern due to competition with a single TF.

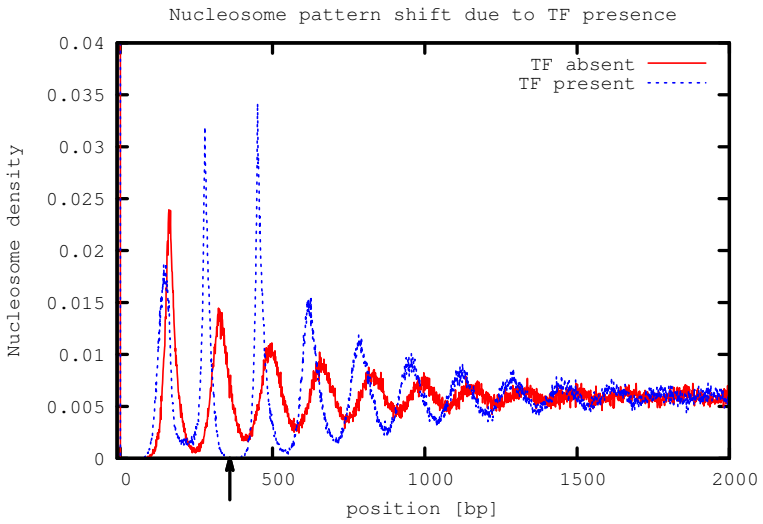


Figure 3.2: Time-averaged nucleosome positioning from stochastic kinetic simulations, assuming a fixed nucleosome at $x = 0$ as well as a TF of size $m = 10$ that binds specifically at $x = 360$ with energetic preference $\mu_{\mathrm{TF}}(360) = \mu_{\mathrm{N}} = 9k_BT$ (the TF binding position is indicated by the black arrow). The difference between the red trace (TFs removed) and the blue dashed trace (where TFs bind preferentially) illustrates the resulting shift in nucleosome patterns from the binding of the TF which, effectively, serves as an additional barrier.

As one can see from Fig. 3.2, the nucleosome density pattern is visibly perturbed over a significant length due to the TF. The new oscillating density pattern includes a new set of loci available for TF binding since the set of sites found in linker regions or peripheral nucleosome (susceptible to exposure from transient unwrapping) has shifted.

3.1.1 Nucleosomal Energetic Profile in the Promoter Region

TFs are, however, not the only factors that bind preferentially at certain target points. Until now we have assumed uniform binding affinity adjacent to a fixed +1 reference nucleosome that acts as a boundary for particles downstream. In Chapter 4 we will explicitly consider the 'landscape' of nucleosome binding energetics. For the moment, however, as a crude approximation to nucleosome binding patterns in the promoter region, we simply assign a position dependent binding affinity μ_{N} with x to capture the global energetics of binding throughout the entire NFR and +1 positioning region, rather than a fixed +1 barrier.

A very simplified profile is shown in Fig. 3.3(a); here we assume the NFR, an area of low binding affinity, has a size of $h = 200$ bp consistent with experimental estimates[66]. The

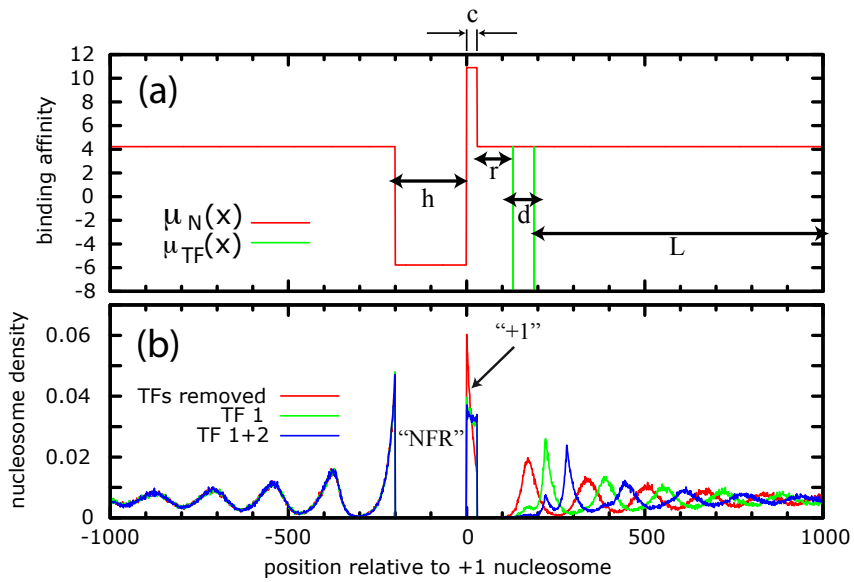


Figure 3.3: (a) sketch of simple profiles for the binding affinity of nucleosomes μ_{N} (red) and of TFs μ_{TF} (green). (b) Nucleosome density pattern in the absence of TFs (red), with a single TF at $x = c+r$(green), and with TFs binding at both $x = c+r$ and $x = c+r+d$ positions indicated in (a).

width of the high-affinity +1 positioning peak was taken to be $c = 30$ bp, consistent with typical convolution ranges of the +1 peak from yeast data in Ref. [36]. Figure 3.3(b) provides the density profile throughout the entire promoter region using this simple position-dependent nucleosome binding affinity. The figure is highly idealized (for example, the near complete elimination of histones from the NFR is exaggerated) but is intended to show the versatility of the numeric approach. Again, the influence on nucleosome density by a transcription factor at $x = 130$ can be seen in the green data. With a second TF at $x = 190$; the resulting blue data shows an even further displaced +2 peak. The model used here assumes no strong 'positioning' of nucleosomes away from the promoter region, and so displacement of nucleosomes by occlusion with TFs is incremental. Nevertheless, the curves in Fig. 3.3 suggest that two or more TFs can increase their effectiveness in displacing a nucleosome by working in concert.

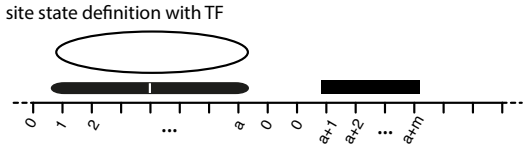


Figure 3.4: Schematic of site state definition analogous to Fig. 2.3 including states corresponding to transcription factor binding at right. Recall that state i is represented in the matrix of Eq. 3.2 by row $i+1$.

3.2 The Two-Factor Transfer Matrix

In addition to the stochastic simulations in the previous section, the displacement of nucleosomes due to competition with TFs can be observed by adapting the transfer matrix approach from Chapter 2. To do so, we define an additional set of m states corresponding to binding positions along a transcription factor m bp in length; this state convention is illustrated in Fig. 3.4. The corresponding transfer matrix T_{NTF} (the subscript 'NTF' denotes 'nucleosomes + transcription factors') for these states is $a+m+1$ in length. As in Chapter 2, we assume w bp are subject to unwrapping in either direction from the nucleosome dyad and a footprint length $a = 2w+1$. T_{NTF} contains the matrix $\tilde{T}_{\mathrm{NN}}$ (containing exclusively nucleosome-nucleosome interactions) given in Eq. 2.9 as a submatrix in the upper left corner:

$$\tilde{T} = \begin{bmatrix} & & & & & & 1 & 0 & \dots & & 0 \\ & & & & & & 0 & & & & \\ & & & \left[\tilde{T}_{\mathrm{NN}}\right] & & & \vdots & \vdots & & & \vdots \\ & & & & & & e^{\mu_N - u(1)} & & & & \\ & & & & & & e^{\mu_N - u(2)} & & & & \\ & & & & & & \vdots & & & & \\ & & & & & & e^{\mu_N - u(w)} & & & & \\ & & & & & & 1 & 0 & \dots & & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & e^{\mu_{TF}} & 0 & \dots & 0 \\ & & & & & & 0 & 0 & 1 & & 0 \\ \vdots & & & & & & & & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & & 0 & & & & 0 & 1 \\ 1 & 1 & e^{-u(w)} & e^{-u(w-1)} & \dots & e^{-u(1)} & 0 & & \dots & & 0 \end{bmatrix} \tag{3.2}$$

where μ_{TF}, and μ_{N} denote the binding energy of TFs and nucleosomes respectively. As mentioned above, TFs target certain sequences for binding quite specifically, implying strong position dependence in the elements of Eq. 3.2 involving μ_{TF}. For this reason, we define a matrix $\tilde{T}_{\mathrm{TF}}$ for specific-binding TF loci where $\mu_{\mathrm{TF}} = \mu_{\mathrm{N}} = 4$, and take the limit $\mu_{\mathrm{TF}} \to \infty$ elsewhere. As in Section 3.1.1, we can incorporate position-dependent features of nucleosome formation by varying μ_N with position, and we define the matrices $\tilde{T}_{\mathrm{NFR}}$, $\tilde{T}_{+1}$, and $\tilde{T}_{\mathrm{NS}}$ to denote the transfer matrices for the nucleosome free region, the +1 binding region, and the non-specific region far away from the promoter respectively (see Fig. 3.3(a)). The partition sum for a system of size L including the full promoter region is then

$$Z_{\mathrm{NTF}} = \mathrm{Tr}\left\{\tilde{T}^{h}_{\mathrm{NFR}} \cdot \tilde{T}^{c}_{+1} \cdot \tilde{T}^{r}_{\mathrm{NS}} \cdot \tilde{T}_{\mathrm{TF}} \cdot \tilde{T}^{d}_{\mathrm{NS}} \cdot \tilde{T}_{\mathrm{TF}} \cdot \tilde{T}^{L}_{\mathrm{NS}}\right\}. \tag{3.3}$$

By using Eq. 3.3 in conjunction with projection matrices and taking the large L limit just as in the derivation of Eq. 2.15, state probabilities for various positions can be calculated; equivalent calculations are provided in full in Ref. [65], but are omitted here to avoid repetition. One benefit of this method is that it allows for efficient calculation of cooperativity in TF binding.

3.3 Transcription Factor Cooperativity

Fig. 3.3 illustrates the combined effect of multiple TFs to cooperatively displace or evict a nucleosome; naturally, this would suggest that the presence of one TF can facilitate the binding of other TFs in the vicinity. Cooperativity between TFs has been observed experimentally *in vitro* [67], and *in vivo* [68], and has been studied theoretically in Ref. [69].

Making use of the formalism from the latter of these studies, two arbitrary sites for TFs a and b are assigned statistical binding weights q_a and q_b respectively. In the absence of any binding at b, the probability that site a is occupied is $\tilde{p}_a = q_a/(1+q_a)$ (and likewise for protein b in the absence of binding by a, $\tilde{p}_b = q_b/(1+q_b)$.) More generally, however, when accounting for possible binding of other proteins and an unknown interaction, the probability of binding at site a, p_a, the probability of binding at site b, p_b, and the probability of binding at both positions simultaneously $p_{a,b}$ are given respectively as[69]

$$\begin{aligned} p_a &= \frac{q_a(1+\omega\cdot q_b)}{1+q_a+q_b+\omega\cdot q_a\cdot q_b} \\ p_b &= \frac{q_b(1+\omega\cdot q_a)}{1+q_a+q_b+\omega\cdot q_a\cdot q_b} \\ p_{a,b} &= \frac{\omega\cdot q_a\cdot q_b}{1+q_a+q_b+\omega\cdot q_a\cdot q_b}. \end{aligned} \tag{3.4}$$

Eq. 3.4 makes use of the dimensionless cooperativity ω which, on a log scale, captures the degree to which the presence of one TF promotes (or perhaps even inhibits) the binding of the other. ω can be solved for as

$$\omega = \frac{(p_a + p_b - p_{a,b} - 1)\, p_{a,b}}{p_a p_{a,b} + p_b p_{a,b} - p_a p_b - p_{a,b}^2}. \tag{3.5}$$

For $\omega > 1$ the two TFs bind cooperatively, while values of $\omega < 1$ imply antagonistic binding. The coupling constant $J_{1,2}$ between the two is then defined as $\omega_{1,2} = e^{-J_{1,2}}$

Obviously this cooperativity will depend on the proximity of loci a and b as well as their relative binding strength and the prevailing pattern of nucleosomes due to other factors. Notably, a and b need not necessarily be directly adjacent to one another. As noted in

Fig.'s 3.2 and 3.3 the presence of one TF on the DNA shifts the pattern of nucleosome density distribution, potentially uncovering or occluding the loci of secondary TFs even when separated by several nucleosomes. Long-range cooperativity between TFs *in vivo* has been pointed out by Vashee *et al*, though via a somewhat different mechanism[70] from what has been presented above.

For concreteness, we return to the example used in Fig. 3.2 except with a nucleosome fixed at $x = 0$ ($c = 0$), and TFs with $m = 5$ located at r and $r + d$ bp downstream. Fig. 3.5 charts the cooperativity $\ln(\omega)$ from Eq. 3.5 between TF a at r and TF b at $r + d$ using the equilibrium calculation based on Eq. 3.3.

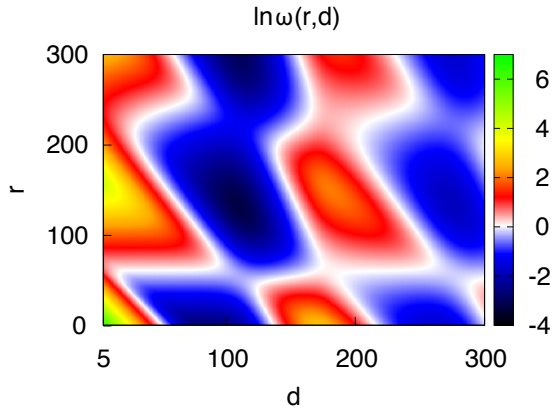


Figure 3.5: Heat map of cooperativity ω from Eq. 3.5 between two TFs separated at a distance d from each other with the first a distance r from the +1 positioned nucleosome. Reproduced from Ref. [65] with permission.

3.4 Experimental Observations

So far, this chapter has been purely theoretical. Our hope, however, is to compare these predictions with biological data. Before describing our efforts in doing so, a small amount of context is necessary. In the yeast *Saccharomyces cerevisiae*, fermentation is the major means of energy production (even in the presence of oxygen). When glucose becomes scarce, a switch is made to respiration using ethanol as a carbon source, resulting in massive reprogramming of gene expression[71, 72] that can hopefully be exploited to analyze changes in chromatin structure.

To attempt to connect the above observations to experimental data, Ref. [62] undertook a review of data on the *Saccharomyces cerevisiae* genome under such changing growth conditions. Available data on nucleosome occupancy[73], as well as expression level[74] of yeast in environments rich in glucose, galactose, and ethanol respectively were analyzed. Particular attention was paid to occupancy in the promoter regions of genes that underwent the greatest changes in expression. The report identified changes in nucleosome occupancy in the promoter region of genes with the strongest relative change in expression and the position of binding sites for TFs related to the above metabolic switch.

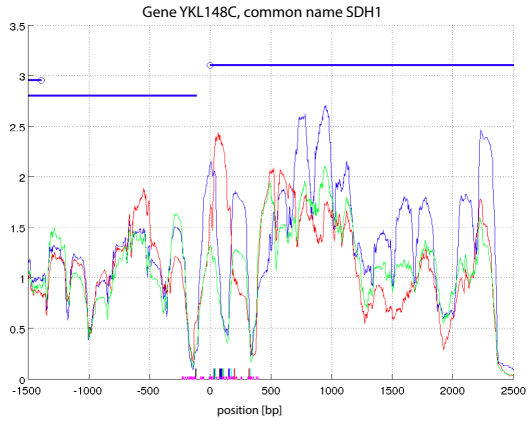


Figure 3.6: Shift in the +1 nucleosome positioning in the promoter region under the influence of TF binding. The blue line above denotes the transcript region with the circle representing the TSS. Occupancy in glucose is shown in blue, in ethanol in red and in galactose in green. TF binding sites are marked at the bottom with HAP1/2/3/4/5 in blue, GAL4 in green, MIG1/2/3 in orange, NHP10 in black, ADR1 in red, and MSN2/4 in cyan. Note the particularly dense concentration of HAP binding sites in the +1 region overlapping with the +1 nucleosome. Reproduced from Ref. [62] with permission.

Many examples of shifted nucleosome patterns in the vicinity of TF loci are listed in Ref.[62], and changes in nucleosome occupancy and TF binding activity showed some consistency with the above picture of competitive binding, however a global rule remained elusive. Several exceptions were observed. For example, the HAP4 TF is induced during the shift from fermentative to respiratory metabolism[75, 71]. Fig. 3.6 shows occupancy changes in the promoter region of the gene YKL148C, with particularly dense HAP binding sites indicated at the bottom of the figure. The shift in nucleosome occupancy at the +1 nucleosome position with ethanol (red) is clearly visible. However, unexpectedly, in the ethanol medium the +1 nucleosome is shifted *towards* the binding sites of the enriched HAP TFs, the opposite of what would be expected from the model of binding occlusion presented earlier.

As a possible explanation for this, homologs of the *S. cerevisiae* Hap(2-5) proteins Hap(A-E) in *Aspergillus nidulans*, have demonstrated bending of DNA similar to that in nucleosomes. They also have the potential for "mutual substitution in nucleosomes and [...] interactions with the histones H3 and H4 in mixed tetrasomes."[76] If such *nucleosome-like* proteins are formed by combining HAP proteins with H3 and H4 histones, then it is possible that this would account for observed occupancy shifts toward overlap with HAP TFs as illustrated in Fig. 3.6. This is, however, quite speculative. At any rate, our observations suggest that a general picture of TF-nucleosome interactions along DNA requires a more subtle description than the simple competitive interaction potential that was posited in Eq. 3.1.

4 The Inverse Problem

Up to now, our consideration of the influences on nucleosome positioning has been mostly restricted to effects arising from competitive binding -either between the nucleosomes themselves, or between nucleosomes and transcription factors- and the statistical positioning that arises from this interaction. In Chapter 2 we also introduced one experimentally-motivated mechanism for remodeller action that was shown to be particularly important at low histone densities.

In addition to these effects, however, another ever-present influence on nucleosome positioning is the inherent binding affinity of different locations along the DNA encoded in its sequence. We refer to this as the 'energy landscape' of the genome: regions where the landscape is low are attractive to nucleosome formation, while regions of high potential are repulsive.

Omission of this effect in Chapter 2 was justified by taking averages over many different genes throughout the yeast genome, with alignment to +1 nucleosome barrier. There, the assumption was that the sequence-encoded features over many genetic loci across the genome would average out to an approximately flat landscape, and that there is insignificant *systematic* trend, genome-wide, in the landscape relative to +1 nucleosomes. Although this latter assumption was not proven, the resulting analysis was predicated to some degree on Occams Razer. As Bertrand Russel put it: "Whenever possible, substitute constructions out of known entities for inferences to unknown entities."[77]. With this advice in mind, we assert that if sequence-dependent landscape features are not *necessary* to reproduce gene-averaged experimental data, then an appropriately parsimonious model should not include them. At the single-gene level, however, landscape features play a much more prominent role.

Thus, we return to this omission and address the effects on nucleosome positioning from the sequence-encoded landscape itself, a subject in which a great deal of work has already been invested. For example, it is known that in *Saccharomyces cerevisiae*, certain sequences of DNA, such as Poly(dA:dT) tracts are associated with nucleosome-depleted regions, perhaps due to their inherent stiffness against bending[24, 78], though this has been shown not to be consistent with other species of yeast, such as *Schizosaccharomyces pombe*[79]. Indeed, there is also evidence supporting the additional stiffness of the AA/TT bond at the bp level while the stiffnesses of other pair bonds seem not to obey any simple rule[80]. This suggests that the effective histone affinity of a given stretch of DNA is actually determined by a complex interplay of many factors.

Kaplan *et al* have developed a computational model to predict histone affinity from an arbitrary sequence across the yeast genome[73], while Gossette and Lieb have shown that under artificially reduced histone conditions, sequence effects become increasingly important in determining which nucleosomes are preserved, and which are not[81]. From a more theoretical perspective, Chereji *et al* have proposed a means for inferring the landscape based on a presumed 10-bp periodicity in linker length[82, 83], while Teif and Rippe have proposed a model

for predicting nucleosome positions based on sequence and remodeling activity[84]. For these and many more reasons, the influence of the energetic landscape cannot be excluded from discussion of nucleosome positioning at the single-gene level.

An important caveat here is that we are referring to an *effective* landscape in the sense that it represents the net influence of all non-translationary-invariant nucleosome positioning effects. The elasticity-dependence of DNA on sequence is a fairly intuitive mechanical dimension of the energy landscape, as discussed above. Another, less obvious, component of the landscape, however, comes in the form of active remodellers that target specific loci for remodelling and repositioning of nucleosomes.[85, 86, 87] The SWI/SNF family of remodellers, for example, tends to target acetylated histones and serves to clear histones from the nucleosome free region. The ISWI family of remodellers, on the other hand, activates 'sliding' at the interface between nucleosomes and linker DNA; ISWI2 activity is focused at the +1 nucleosome while ISWI1 tends to have greater influence on the spacing of the nucleosomes downstream[25]. Naturally, then, changes in remodelling activity lead to altered nucleosome positioning that will be reflected in the effective landscape of a given region of DNA. The following calculation of this landscape is not intended to evince the precise mechanism responsible for preferred nucleosome positioning, but rather to capture the global effective specificity of nucleosome positioning arising –directly or indirectly– from the genetic sequence, all things considered. One may hope that mechanistic understanding can then be gleaned from controlled changes in the conditions under which the landscape is determined.

4.1 Introducing the Potential Landscape

For explicit inclusion of sequence-dependence in our lattice model, we consider 4 related quantities which can be subdivided into analogous pairs. The energetic landscape, $V(x)$, is the focus of this chapter and captures the potential energy inherent in a one-dimensional particle[1] being bound to a substrate at position x in isolation –i.e. without any energetic costs incurred from neighboring particles. The latter are explicitly included in the two-body interaction potential $\phi(x, x')$ between a pair of neighboring particles at x and $x' > x$.

The density of particles at a position x is defined as $n_1(x)$, while the two-body pair density $n_2(x, x')$ describes the density of nearest-neighbor pairs at x and x'. The analogy between the two pairs of quantities $[n_1(x), n_2(x, x')]$ and $[V(x), \phi(x, x')]$ relates to one- and two-body effects and either pair can be inferred from the other, as will be shown below. To do so however, we must take note of two simplifications implicit in this starting point.

First, there is, in principle, no reason *a priori* why we must restrict ourselves to one- and two-body effects. For example, a so-called 'three body' potential $\phi_3(x, x', x'')$ and 'triplet density' $n_3(x, x', x'')$ could be introduced to represent nucleosome clustering due to, for example, the influences of higher-order chromatin structure that extend beyond nearest-neighbors. Similar higher-order interactions and distributions could be extended to arbitrarily high order, with corresponding complexity. Restricting nucleosome interactions to nearest-neighbors, however,

[1] throughout this chapter, we will refer more generally to 'particles' in one-dimension –nucleosomes along DNA being one such particle

is in line with the mechanistic assumptions about neighbor interactions made in previous chapters, and renders the problem analytically tractable.

Secondly, it is conceivable that the neighbor interaction ϕ could itself exhibit some position dependence that is not captured purely by the difference in position between the neighboring particles. Based on the physical argumentation from Chapter 2, however, we assume a translationally invariant competitive binding interaction, with unwrapping cost ε per bp

$$\phi(x, x') = \phi(x' - x) = \phi(\Delta x). \tag{4.1}$$

For the form of ϕ in Eq. 4.1, later in this chapter we will simply take $\phi(\Delta x) = v(\Delta x)$ from Chapter 2, but for the moment we will use ϕ to underscore the generality of the following argument. To some degree, both of these simplifications are, again, motivated by the fact that $V(x)$ and $\phi(\Delta x)$ represent 'known constructions', as there is no apparent need to invoke a more complex set of interactions.

4.2 The 1-D Inverse Problem

Converting between 1- and 2- body particle densities and 1- and 2- body energetic potentials for nearest-neighbor interacting particles along a 1-D axis has been considered[88, 89, 90]. This work by Percus provides the fundamental tools for the calculations in this chapter; the derivation below from Eq.s 4.2 through 4.15 represents recapitulation of the relevant aspects of this work, with some added commentary and a minor correction midway, though the conclusion is identical. We follow this reasoning explicitly since an understanding of this methodology is necessary for the application that will follow.

For a 1-D lattice of size L, we assume a corresponding basis set of vectors $|i\rangle$ (with length L) that denote the position of particle x_i. That is to say, if particle x_n is at position m, then the m-th elements of $|n\rangle$ is 1 and all others are zero. This leads to the unconstrained boundary condition vector

$$|J\rangle = \sum_{n=0}^{L} |n\rangle. \tag{4.2}$$

Hence, if all particle interactions and positions were energetically neutral, our partition function for a single particle would be simply $Z = \langle J|\, J\rangle$. For a particular particle (say, the i-th particle $\langle i|\, J\rangle = 1$) We define $\phi(1,2) = \langle 1|\phi|2\rangle$ to be the energy of interaction between the two particles $x_1 < x_2$ and thus

$$\langle 1|w|2\rangle = \mathrm{e}^{[-\beta\phi(1,2)]}\Theta(x_2 - x_1) \tag{4.3}$$

where Θ is the heaviside function, and β is inverse temperature. Again, our application allows for the more restricted $\phi(1,2) \to \phi(\Delta_{1,2})$, but the more general form is left in place. The diagonal matrix e contains the weights of particle positions in isolation

$$\langle 1|e|2\rangle = \mathrm{e}^{[-\beta V(x_1)]}\delta(x_2 - x_1) \tag{4.4}$$

where the diagonal elements of e are simply the Boltzmann factors of $V(x)$. Assuming a given particle number N, we define a canonical partition function Q_N. For $N = 1$, the lone particle has no neighbors with which to interact, and thus a partition sum of

$$Q_1 = \langle J|\, e\, |J\rangle\,, \tag{4.5}$$

accounting for all possible positions of the particle. Naturally then, the partition sum for two particles involves two Boltzmann weights for specific binding in addition to a single interaction matrix

$$Q_2 = \langle J|\, ewe\, |J\rangle\,. \tag{4.6}$$

Likewise, for arbitrary N,

$$Q_N = \langle J|\, (ew)^{N-1} e\, |J\rangle\,, \tag{4.7}$$

Assuming $N > 0$, otherwise $Q_0 = 1$ corresponding to the empty state.

In the *grand canonical* ensemble, however, we must sum over the set of all possible numbers of particles N. To accomplish this, we augment the Boltzmann matrix of specific binding with an overall chemical potential μ that represents the basic non-specific binding energy for nucleosomes to be formed *anywhere*. Hence, we replace the previous matrix e with $z = \mathrm{e}^{[\beta\mu]}e$, yielding:

$$\langle 1|\, z\, |2\rangle = \mathrm{e}^{[\beta(\mu - V(x_1))]}\delta(x_2 - x_1). \tag{4.8}$$

The grand canonical partition function is then:

$$Z = \sum_{N=0}^{L} \mathrm{e}^{[\beta N\mu]} Q_N = 1 + \sum_{N=0}^{L} \langle J|\, (zw)^{N-1} z\, |J\rangle\,. \tag{4.9}$$

At this point, we exploit the matrix generalization of the geometric series (or, the Neumann series)

$$\sum_{k=0}^{n} A^k = (I - A)^{-1}(I - A^{n+1}), \tag{4.10}$$

assuming $(I - A)$ is invertible, and the eigenvalues of A are less than unity. In our case, $k \to N - 1, A \to zw$, and $A^L \to 0$ for large L (i.e. the probability of having an entire array stacked with particles *everywhere* is vanishingly small) and truncation artifacts of this series can be neglected. Hence

$$Z = 1 + \sum_{N=1}^{L} \langle J|\, (zw)^{N-1} z\, |J\rangle = 1 + \langle J|\, (I - zw)^{-1} z\, |J\rangle\,, \tag{4.11}$$

which differs slightly from Eq. 2.6 of Ref. [90]. By extending this argument in either direction from a designated particle, the 1−particle density distribution $n(1)$, is seen to be:

$$\begin{aligned} n(1)Z &= \langle J| (I - zw)^{-1} z |1\rangle \langle 1| z(I - wz)^{-1} |J\rangle \\ &= \langle J| (I - zw)^{-1} |1\rangle z(x_1) \langle 1| (I - wz)^{-1} |J\rangle \end{aligned} \tag{4.12}$$

Likewise, the 2−particle neighbor function is given by

$$n_2(1,2)Z = \langle J| (I - zw)^{-1} |1\rangle z(x_1) \langle 1| w |2\rangle z(x_2) \langle 2| (I - wz)^{-1} |J\rangle . \tag{4.13}$$

Eq.'s 4.12 and 4.13 suffice to carry out what could be termed the 'forward' calculation. That is to say, the determination of the 'effects' ($n_1(x)$, and $n_2(x, x')$) from the 'causes' ($V(x)$ and $\phi(x, x')$). The above description also suffices to present a qualitatively clear understanding of the relevance of this work. The remaining calculations to derive the analogous 'inverse' calculations (i.e. inferring the energetic interactions based on the observed 1- and 2-body density patterns) can be found in the original Ref. [90]. The results of this calculation, are:

$$\begin{aligned} \beta [u(x_1) - \mu] &= \ln \left[1 - \langle J| (I - n_2 n^{-1}) n |J\rangle\right] - \ln \left[\langle J| (I - n_2 n^{-1}) |J\rangle\right] \\ &- \ln [n(x_1)] - \ln \left[\langle 1| (I - n^{-1} n_2) |J\rangle\right] \end{aligned} \tag{4.14}$$

$$\begin{aligned} \beta\phi(x_1, x_2) &= \ln \left[\langle 1| (I - n^{-1} n_2) |J\rangle\right] + \ln \left[\langle J| (I - n_2 n^{-1}) |2\rangle\right] \\ &- \ln \left[\langle 1| (n^{-1} n_2 n^{-1}) |J\rangle\right] - \ln \left[1 - \langle J| (I - n_2 n^{-1}) n |J\rangle\right] \end{aligned} \tag{4.15}$$

Eq.'s 4.12–4.15 define the forward and reverse calculations for the two pairs of quantities discussed above. In principle, they sufficiently constrain the landscape and two-body interaction potential provided the full statistics of particle positioning can be obtained. However, for nucleosomes, such statistical data is often incomplete in practice, and in the following section we explore how the above analysis can be practically applied to nucleosome data.

4.3 Application to Nucleosome Positioning Data

The preceding has shown the feasibility of interchangeably calculating the pair of 1- and 2-body energetic quantities from corresponding density quantities and vice-versa. Given the availability of MNase positioning data, and the neighbor interactions described in previous chapters, it may seem that the remaining quantities of interest can be easily calculated.

However, while experimental techniques provide n_1 up to a normalization factor, the process of read-collection omits all information related to $n_2(x, x')$. Since many cells are used in the process, it is impossible to differentiate which mononucleosomes had previously been adjacent to one-another on the same genome. Conversely, the theoretical discussion from Chapter 2 provides a description of neighbor-interactions $\phi(x) \to v(x)$, but omits any consideration of

the potential landscape, V. Thus, both the forward and reverse calculations above are missing one half of the necessary input.

In the face of this conundrum, an iterative scheme to calculate V was developed. This scheme relies on taking the theoretical description of ϕ given in Chapter 2 (i.e. the SoNG potential $v(\Delta x)$) along with successive candidate functions for V nominated iteratively to calculate n_1 theoretically until deviation from experimental data is minimized.

In this scheme then, the landscape value $V(x)$ is a free parameter at every value of x, and convergence at a solution for $V(x)$ requires optimization over an L-dimensional parameter space. For a typical system of L in the thousands, an exhaustive mesh comparison over all $V(x)$ values would be computationally infeasible, and so an optimized search algorithm is required.

High-dimensional numeric optimization is a field that encompasses many different techniques; one of the most standard[91] of these is the Nelder-Mead, or 'amoeba' search algorithm[92]. This method has been adopted for high-dimensional numeric optimization successfully in a variety of fields. Its power lies in its ability to infer regional trends in the function's slope from sample evaluations without needing to explicitly calculate the gradient of the function at any particular point. For these reasons, it is well-suited to our goal of determining $V(x)$.

4.3.1 The Amoeba Method

We seek a numeric scheme to numerically minimize a functional f of the euclidean distance between the nucleosome read counts observed experimentally $\Omega(x)$, and the density predicted theoretically from Eq. 4.12, $n_1(x)$, with optimization over the set of parameters $V(x)$ for all positions x. We define the functional subject to minimization explicitly as:

$$f([V(x)], \phi(\Delta x), \Omega(x)) = \sum_x [\Omega(x) - \alpha n_1(x)]^2, \tag{4.16}$$

where α is an unknown normalization constant associated with the number of cells sampled in the experiment, and is assigned analytically

$$\alpha = \left(\sum_x n_1(x)\Omega(x) \right) / \left(\sum_x n_1(x)^2 \right) \tag{4.17}$$

(See Ref. [36], SI). We assume that f is a smoothly-varying function in the space of $V(x)$, and while $n_1(x)$ is written merely as a function of x, dependence on V and ϕ, as described above, is assumed. Clearly, f depends on the experimental read counts $\Omega(x)$, as well as the set of parameters $V(x)$ to be optimized, and the neighbor interaction $\phi(\Delta x)$ which we assume to be the SoNG potential $v(\Delta x)$ from Chapter 2.

To implement the amoeba search algorithm, a set of $L + 1$ candidate functions $V_i(x)$ are generated (each of which correspond to their own unique energetic landscape), representing

vector positions throughout this $L-$dimensional parameter space[2], and are each considered as possible solutions for optimization. A convenient choice for initial candidates is given by

$$V_i(x) = \begin{cases} 0 \,\forall\, x & \text{if } i = 0 \\ \pm 1 \cdot \xi \, \delta_{ix} & \text{if } i \in [1, L] \end{cases} . \tag{4.18}$$

In Eq. 4.18, the factor ξ represents an intuitive guess at the characteristic length scale over which $n_1(x)$ changes significantly with $V(x)$, and can be arbitrarily assigned a value of $1k_BT$ without significantly affecting the final results. With V_0 representing a baseline 'flat' energetic landscape, each other candidate landscape is initially shifted by either $+\xi$ or $-\xi$ (randomly, with equal probability) from zero at a single position. $\overline{V}$ is the centroid of these points in the L-dimensional space, and the L vectors $V_i - V_0$ form a linearly independent basis, the correct combination of which yields the point in the space which minimizes the functional f.

To determine that combination correctly, the initial candidate points V_i are ordered by their corresponding f_i values. V_l is then the point with the lowest f value, while V_h corresponds to the highest, and is iteratively replaced with transformed positions to reduce its f value. These transformations are described in more detail using the flow-chart in Appendix A.2, Fig. A.1 and are repeated until eventually all candidate points converge to the same position in parameter space with the same f value within numeric tolerance. A full description of this algorithm as well as the criteria for convergence can be found in Ref. [92].

4.3.2 An Illustrative Minimal Example

Even with the description in A.2, and the accompanying diagram, Fig. A.1, the application of the amoeba method to our landscape problem may still seem unintuitive. For this reason, it is helpful to consider a minimal illustrative example where $L = 2$ –an extremely simplified scenario.

In this case, the two degrees of freedom to the potential V are denoted $V(x)$, and $V(y)$ respectively, as shown in Fig. 4.1, and within this space, the optimum solution which leads to the correct density exactly **E** is sought (E for 'exact' is a useful mnemonic). To that end, three initial guesses are made in the 2-D parameter space of V_x, V_y in Fig. 4.1(b): points **A**, **B**, and **C**, color-coded red, blue and green in the figure respectively. Each of these points, however, also represents a candidate landscape sketched in Fig. 4.1(a). In this figure, **A** represents V_0 from Eq. 4.18 and is the default, or 'flat', featureless landscape, while **B** and **C** contain random linear displacements from **A** at positions x and y respectively. Since **C** has the highest scalar functional value in Fig. 4.1(b) it is reflected about the centroid to point **C***. Thereafter, **B** has the next poorest fit and will also be substituted. The algorithm will continue to make iterative substitutions until all candidates converge at **E**.

Obviously, calculating a genomic landscape requires considering far more than $L = 2$ adjacent positions (typical genomic segments under consideration range in sizes of approximately $L \approx 10,000$) and the dimensionality of such parameter space searches is not conducive to visualization. Nevertheless, the same principle applies: the set of $L + 1$ vertices form a simplex

[2] bp are treated as lattice sites here, ignoring the option of coarse-graining

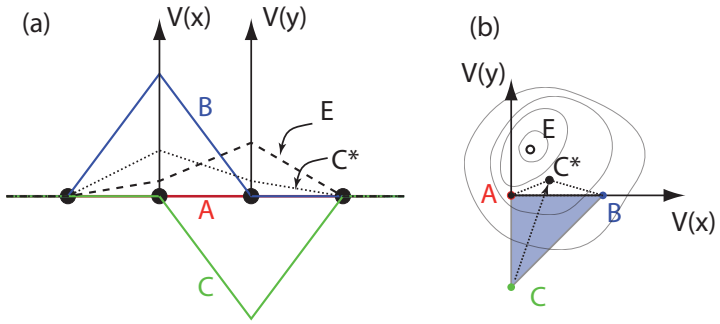


Figure 4.1: Schematic of a highly reduced inverse-landscape algorithm with only two values $V(x)$ and $V(y)$. (a), Various potential landscapes corresponding to the potential at these two positions are sketched: an initially flat landscape, **A**, along with two landscapes with random displacements at the two positions, **B**, and **C**. The optimal, or 'Exact' solution is the dashed line **E**, and two axes 'V(x)' and 'V(y)' indicate the degrees of freedom open to the system. All of these profiles correspond to a point in the 2-D plane in figure (b). Here, a simplex of $N + 1 = 3$ points is shown, lines of constant f are sketched in transparency. Points are drawn only approximately to scale. Since **C** has the highest f value, it is substituted with **C*** before the fitness of each points is reevaluated; upon the next iteration, **B** will be substituted for a new point **B***. Eventually, all points will converge to the exact solution **E** within numeric tolerance.

in the $L-$dimensional space which iteratively converge to a single position representing the global minimization in f.

As a proof of this principle, an intermediate test case is proposed in which the exact solution **E** is known, but 'hidden' from the algorithm, along a segment for $L = 100$. For this confirmation see appendix Fig. A.2.

4.3.3 The 601 Sequence

We may now proceed with application of this algorithm to actual experimental data using a sequence known particularly for its high-histone affinity[93, 94], which has come to be referred to as the 601 sequence. Nucleosomes reconstituted *in vitro* on plasmid DNA segments with histones and DNA mixed in mass ratio 1:1 and 0.5:1 respectively were sequenced, and the resulting positioning data was made available for the collaborative analysis above[95]. Naturally, the overall densities of nucleosome positioning differed, depending on histone mass concentration. However, assuming the above analysis is valid, the underlying potential landscape ought to be independent of histone density. For this application, two interactions $\phi(\Delta x)$were considered: (1) neighboring nucleosome interactions exactly as obtained in the SoNG model, as described in Eq. 2.5 from Chapter 2, and (2) the neighbor-potential from Eq. 2.5 in addition to a slight linear attractive component to account for possible clustering due to higher-order chromatin structure or histone-tail interactions. The results of the fits are shown in Fig. 4.2.

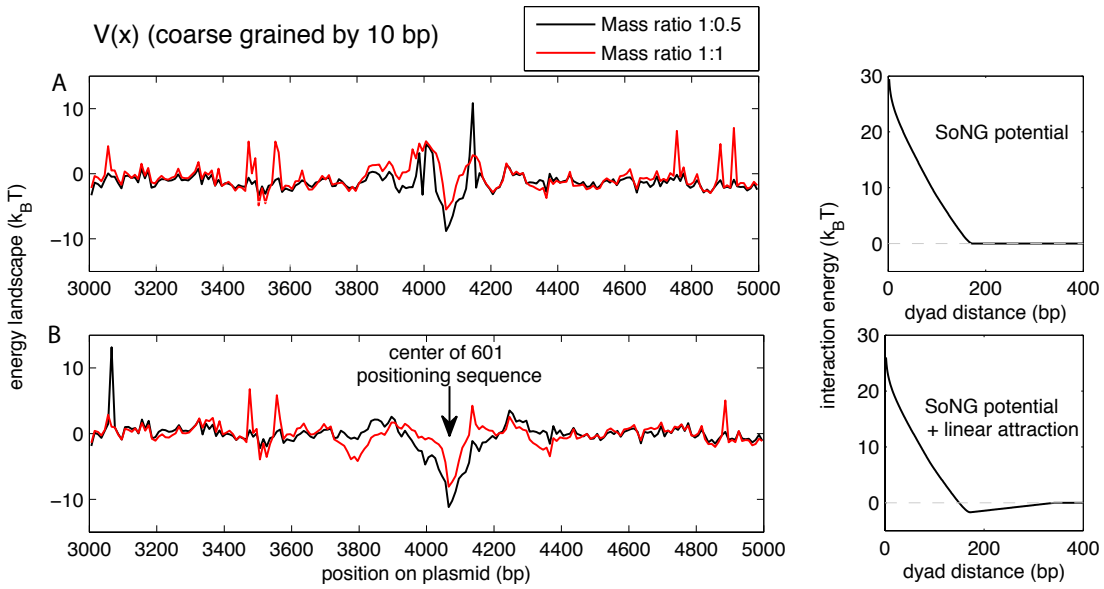


Figure 4.2: Landscape potential inferred from plasmid 601 sequence at different histone concentrations assuming two different neighbor interactions: (A) the SoNG model alone, above, as well as (B) the SoNG interaction in addition to a slight attractive potential to capture possible clustering effects, below. Both potentials are sketched to the right.

4.4 Discussion

It bears repeating that this landscape is an effective one: it incorporates direct effects from DNA due to, e.g. elasticity, as well as indirect specific sequence effects from targeted remodelling when such factors are present. In the data used in Fig. 4.2 there were no remodellers acting on the substrate, however analogous experiments were carried out with remodellers re-introduced and resulting density profiles were compared to the previous as a control. Since different classes of remodellers are known to target different segments of the gene[25], it was proposed that the change in density patterns that can be observed upon addition of different remodellers[95] could be translated to an effective energetic potential -or gradient thereof- in the various segments of the DNA.

Although characteristic changes in density patterns were clearly observed under the influence of remodellers, no additive potential could be associated with each, and further study is needed to make definitive claims about the degree to which $V(x)$ is determined from active remodelling.

One avenue for further development of this technique, however, would be the change in the effective nucleosome potential under changing cell conditions that affect the binding of Transcription factors at known loci. As observed in Chapter 3, competing proteins such as specifically-binding transcription factors influence the binding of histones, and such an effect would be included in the measure of the energetic nucleosome landscape. Such a change in the effective potential could be attractive or repulsive, and the relative change could serve as an indication of relative changes in TF binding activity at particular loci under changing

conditions.

5 Dynamic Processes

Thus far, we have focused on steady state properties of nucleosomes. Chromatin is, however, subject to constant change from the influence of external conditions, active mechanisms, and the periodic need for repair and replication[96]. Nucleosomes can be repositioned[97], or even undergo redistribution of histone proteins between themselves[98]. Nucleosomes can also serve as fluctuating barriers to RNA polymerase that must be circumvented during transcription [99]. One possible mechanism to bypass nucleosomes is via the translocation of loops of DNA around the histone surface[100], which would account for upstream shifts in nucleosome positioning during transcription[17]. Replication, however, is perhaps the strongest example of displacement from equilibrium, during which the amount of DNA is doubled and must be once again bundled into nucleosomes. How the cell is able to quickly restore the high physiological density of nucleosomes in a structured array throughout the genome is not a trivial question. In fact, this problem bears a striking resemblance to a well-studied problem in statistical physics.

The so called 'car parking' problem refers to the random sequential placement of cars, with length unity, along an infinite, unmarked, initially empty street. In this model, cars park on the street with rate r_+ per unit length, and leave with rate r_-; with this intuitive analogy, the adsorption of finite length particles to a 1-D substrate is perhaps more easily visualized. The regime of greatest interest (from a kinetic physics perspective) occurs when $r_+ \gg r_-$, and the formation of regular structured arrays becomes a many-body problem, as many inter-particle gaps must be consolidated to make room for an additional car/particle. Again, this is particularly relevant to the formation of dense nucleosome arrays in yeast where the equilibrium coverage fraction of the genome is estimated at nearly 90% [101] in transcribing regions.

One ostensible shortcoming of this analogy, however, is the assumption of a 'street' that is initially *completely* empty. A eukaryotic genome, by contrast, is never completely depleted of histones, but rather is divided at various segments progressively in a complex process described in Ref. [96]. The assumption of an initially empty substrate, however, becomes insignificant when one considers that the density of adsorbing, mutually-exclusive particles along a 1-D substrate proceeds through a set of characteristic stages. The first of these stages (at low initial particle concentration) is referred to as random sequential adsorption (RSA), and is characterized by rapid density increase until reaching a characteristic 'jamming' density at a coverage fraction of $0.7475\ldots$ [102, 103]. It is at this point, statistically, that remaining spaces are too small to accommodate additional particles. Hence, provided long stretches of DNA are at some point reduced to randomly positioned nucleosomes with concentration below the jamming threshold, the proverb 'all roads lead to Rome', would seem to be a concise description of how ensuing RSA will still lead to the jammed configuration (in the manuscript attached below, we also consider *non-random* initial placement of nucleosomes,

an alternative means of circumventing jamming.) The resulting density and gap distribution of this jammed state has been well characterized[103, 104, 105], and further increases in density require desorption and rearrangement of many neighboring particles, a process which is kinetically very slow.

The biological implications of this fact have been pointed out and simulated in theoretical work [106]. In that work, the kinetics of hard, mutually-exclusive footprints 147 bp in length were simulated as they underwent reversible binding in random positions with rates determined from experiment previously[107]. In that study, it was shown that with random, untargetted binding, a jammed configuration of hard-exclusive nucleosomes along a long stretch of DNA cannot be resolved into an extended array with physiological densities on time scales compatible with the yeast cell cycle. For this reason, the authors proposed lateral positioning remodelers that actively 'slide' nucleosomes until they contact a neighbor. With sufficiently fast remodeller activity, it was shown that high density arrays can be reestablished quickly enough to be compatible with the cell cycle –as they must. Later work on this topic also included the influence of sequence-preference on kinetics[108]. It is, however, natural to wonder whether these are the only explanations.

In this chapter, we relax the idealized 'hard' interaction of neighboring nucleosomes that has been used up to now in modeling nucleosome kinetics. The significance of softness to the underlying biophysics of equilibrium patterns was demonstrated in Chapter 2, however it is in the context of dynamic array-assembly that the nucleosome's soft footprint has the most dramatic implications.

5.1 Nucleosome Cramming

At first glance, it may seem obvious, perhaps even trivial, that introducing softness to one-dimensional particles eases filling and relaxes the jamming phenomenon described above. Gaps that are marginally smaller than a particle footprint admit adsorption, albeit somewhat slower, due to the energetic cost of interaction. It is somewhat less obvious, however, to compare the hard 147 bp HaNG model against the larger 165 bp SoNG model, assuming that the same equilibrium density must be achieved. Also not obvious is the dramatic speed-up the SoNG interaction causes to the equilibration process; we will see later in this chapter that equilibration times between the two models can be separated by orders of magnitude. Finally, and certainly not obvious at all, soft interactions allow for the possibility of a transient maximum in density before *decreasing* toward equilibrium, an observation that will be explored later in the chapter.

For the moment, we begin with a description of the jammed state that occurs at the end of RSA for hard particles of length unity. Up to a mean-field description, the density of voids of size x, $V(x)$ separating common points on hard particles can be extrapolated from published work[105] as:

$$V_{\text{jam}}(x) = \begin{cases} 2\int_0^\infty t\, e^{[-(x-1)t - 2\int_0^t \frac{1-e^{-u}}{u} du]} dt & \text{if } 1 \leq x < 2 \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Note that Eq. 5.1, refers to lengths between common points on each particle, arbitrarily taken at the center, contrary to a common convention of measuring gap size from the inner edge of particles on either side. Here, lengths shorter than unity would entail overlap and are therefore forbidden, while lengths greater than 2 will have already been filled during RSA; the remaining lengths consist only of a band in between. Referring to 'mean-field' in this context, and throughout this chapter, implies the assumption that the probability of a given void having a certain size is independent of the size of the neighboring voids.

Assuming the unit length of Eq. 5.1 actually corresponds to the 147 bp footprint of the HaNG model, and applying the same stochastic simulations described in Section 2.5 (with nucleosome 'particles' adsorbing and desorbing with rates r_+ and r_- respectively) jamming indeed occurs at an average spacing of around 147/0.7475 =197bp, or equivalently, an average density (the inverse of the average spacing) of $5 \times 10^{-3}\text{bp}^{-1}$, as seen in Fig. 5.1. This density is well below the physiological equilibrium density $\rho_{\text{eq}} \approx [1/165]\text{bp}^{-1}$ of typical yeast species, and as mentioned above, reaching such densities via random readsorption is impractically slow for biological purposes[106]. Nevertheless, this jammed configuration serves as an illustrative template for the transition between RSA and high-density equilibration.

For example, when the larger footprint of the SoNG model (165 bp) is considered, we make the assumption that non-overlapping adsorption reactions occur significantly faster than reactions that would require energetically costly overlap with their neighbors. Given this assumption, RSA will lead, at least briefly, to the same jammed configuration described in Eq. 5.1 with the length unit now corresponding to a nucleosome footprint of 165bp, and a jamming density of $165/0.7475{=}223\text{bp}^{-1}$ –even *further* away from equilibrium density than in the jammed HaNG model. In this case, however, the possibility for nucleosomes to 'cram' into binding positions too narrow for full wrapping allows for equilibrium densities to be obtained much faster -on time scales up to 2 orders of magnitude shorter, in fact, as illustrated in Fig. 5.1 for three different values of r_+/r_-. The three different on-rates in Fig. 5.1 were chosen to correspond to three different equilibrium densities near $1/165\text{bp}^{-1}$, given a constant r_- used as a time unit. Real time units at the top of the figure are intended to denote order of magnitude estimates from *in vitro* measurements[107].

What is clear from Fig. 5.1 is that a new phase of the filling process is introduced in between the free filling of RSA and the equilibration period of collective rearrangements for $t > 1/r_-$, and it is this phase in the upper left corner of panel B in Fig. 5.1 which we refer to as 'cramming'. Although the three different on-rates traced in Fig. 5.1 result in jamming at different points in time, it turns out that the void distribution at the point of jamming is the same (see Ref. [109] Fig. S3 attached below). Consistent with the description from Eq. 5.1, the distribution is primarily restricted to a certain set of lengths neither so short as to imply overlap nor so long as to have been filled already. The distribution can be described phenomenologically as

$$G_c(x) = A\,\mathrm{e}^{f(x)} \tag{5.2}$$

where

$$f(x) \;=\; \begin{cases} -v(x) & \text{if } x < a \\ \alpha_1 x + b_1 & \text{if } a \le x < x^* \\ \alpha_2 x + b_2 & \text{if } x^* \le x \end{cases},$$

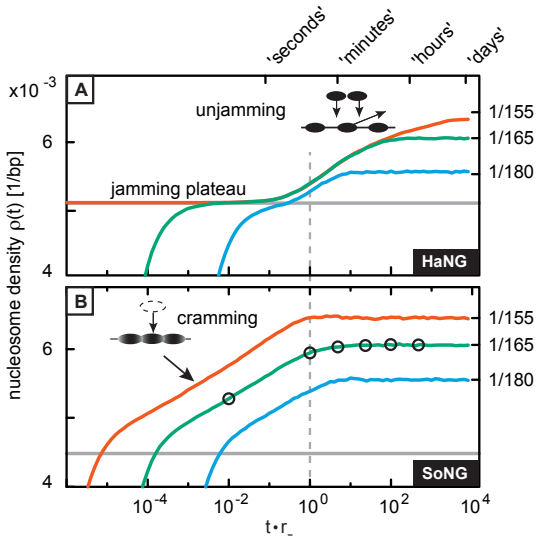


Figure 5.1: Qualitatively different filling behavior can be observed in logarithmic time depending on the neighbor interaction between adjacent nucleosome models; equilibration happens on a much shorter time scale for the SoNG model than for HaNG. Here, the off-rate r_- is set to unity as a reference time unit, while the on-rate r_+ is adjusted to obtain three steady-state densities for both cases: $1/180\text{bp}^{-1}$, $1/165\text{bp}^{-1}$, and $1/155\text{bp}^{-1}$. Time scales above represent only order of magnitude estimates from *in vitro* measurements[107]

and the constants $\alpha_{1,2}, b_{1,2}, A$ and x^* are fit to numerical data with least-squares (see Ref.[109], supplementary information, attached below). A plot of $G_c(x)$ is shown in Fig. 5.2A for the central trace with $\rho_{\text{eq}} = 1/165\text{bp}^{-1}$.

As was the case in the HaNG model, the nucleosomes that make up this distribution remain essentially frozen in place until desorption begins to take effect on time scales of $1/r_-$. Unlike in the HaNG model, however, on time scales faster than that, inter-particle gaps of size x fill with a rate $\gamma(x)$ given by the sum of binding rates at all positions between the two neighbors:

$$\gamma(x) = r_+ \sum_{x'=1}^{x-1} e^{[-v(x')-v(x-x')+v(x)]/k_BT}, \tag{5.3}$$

where $v(x)$ is the energetic interaction between neighboring nucleosomes separated by a distance x as derived in Chapter 2. Eq. 5.3 represents the sum of Boltzmann factors for binding at position x' throughout the void x. The rate is modulated by repulsive interactions on both sides $v(x')$ and $v(x-x')$, while the relieved interaction $v(x)$ is typically negligible (i.e. particles that already overlap are exceedingly unlikely to be filled by a third particle in between). Using Eq. 5.3, it is possible to project how the density might evolve during the cramming phase by considering the sum of explicit filling rates of each of the voids in the system, multiplied by their prevalence; doing so yields

$$\rho(t) = \rho_c + \sum_x G_c(x) \left(1 - e^{-\gamma(x)(t-t_c)}\right) . \tag{5.4}$$

This analytic cramming density as a function of time is illustrated in Fig. 5.2B, with clear correspondence to the data obtaremodelerined directly from our stochastic simulations. Be-

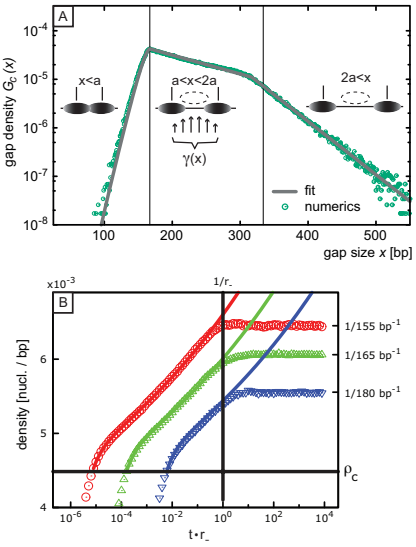


Figure 5.2: A) The actual distribution of distances between neighbors observed in numeric simulations when the jamming density is reached, superimposed is the best-fit function from Eq. 5.2. (B) The distribution from (A) serves as a template for cramming, leading to density described by Eq. 5.4, again superimposed on stochastic data, with evident consistency throughout the cramming period $t \in [1/r_+, 1/r_-]$.

yond time scales of $1/r_-$, equilibration proceeds very quickly, since readsorptions are guided by their neighbor interactions –this effect is discussed in more detail later in the chapter.

Although this model is motivated by physically observed mechanisms, as with the remodelling scheme provided by Ranjith *et al*, it is only one possible explanation for fast array assembly; in practice, chromatin assembly could be influenced by a number of other factors. For example, the processive movement of the replication fork (which, under normal growth conditions in yeast, moves at an average rate of 3kb per minute[110]), given an appropriate speed, can also form nucleosomal arrays quickly without jamming, as gaps are filled immediately behind the replication machinery as soon as they are large enough. As a second example, inherited placement of paternal histones on the daughter chromatids can also serve to prevent jamming, provided the paternal histones are placed at alternating positions on the two daughter chromosomes. This allows for isolated gaps that are each the appropriate size for a single nucleosome. Both of these cases represent a form of non-random initial positioning that avoids the jamming seen in RSA, and thereby alters the filling dynamics.

The expected jam-resolving influence of these effects is demonstrated by theoretical simulation in the published work attached below [109], and, along with the cramming results presented explicitly above, comprise the major findings. A more comprehensive set of computational details and incidental findings have been omitted above for the sake of brevity, but are included in the full report below.

Nucleic Acids Research Advance Access published November 26, 2014

doi: 10.1093/nar/gku1190

Replication-guided nucleosome packing and nucleosome breathing expedite the formation of dense arrays

Brendan Osberg[1,†], Johannes Nuebler[1,†], Philipp Korber[2] and Ulrich Gerland[1,*]

[1]Theory of Complex Biosystems, Physik-Department, Technische Universität München, James-Franck-Strasse 1, D-85748 Garching, Germany and [2]Adolf-Butenandt-Institut, University of Munich, Schillerstrasse 44, 80336 Munich, Germany

Received August 19, 2014; Revised October 30, 2014; Accepted November 3, 2014

ABSTRACT

The first level of genome packaging in eukaryotic cells involves the formation of dense nucleosome arrays, with DNA coverage near 90% in yeasts. How cells achieve such high coverage within a short time, e.g. after DNA replication, remains poorly understood. It is known that random sequential adsorption of impenetrable particles on a line reaches high density extremely slowly, due to a jamming phenomenon. The nucleosome-shifting action of remodeling enzymes has been proposed as a mechanism to resolve such jams. Here, we suggest two biophysical mechanisms which assist rapid filling of DNA with nucleosomes, and we quantitatively characterize these mechanisms within mathematical models. First, we show that the 'softness' of nucleosomes, due to nucleosome breathing and stepwise nucleosome assembly, significantly alters the filling behavior, speeding up the process relative to 'hard' particles with fixed, mutually exclusive DNA footprints. Second, we explore model scenarios in which the progression of the replication fork could eliminate nucleosome jamming, either by rapid filling in its wake or via memory of the parental nucleosome positions. Taken together, our results suggest that biophysical effects promote rapid nucleosome filling, making the reassembly of densely packed nucleosomes after DNA replication a simpler task for cells than was previously thought.

INTRODUCTION

In eukaryotic cells, DNA is packaged into chromatin with nucleosomes as the basic building blocks. A high nucleosome coverage is essential for cells, for example to prevent cryptic transcription (1). In addition, the local positions of specific nucleosomes, especially in promoter regions, can affect transcription factor binding and thereby play an important role in gene regulation (2–5). Nucleosomes consist of about 147 bp of DNA wound around an octamer of histone proteins. While the length of linker DNA connecting neighboring nucleosomes varies locally, nucleosome mapping experiments (6–8) indicate an overall nucleosome coverage of around 90% in yeasts (i.e. the fraction of base pairs of the genomic DNA that are nucleosomal). This dense packing of nucleosomes has to be re-established whenever the DNA is (partially) cleared of nucleosomes, for instance during transcription, repair and replication. This is particularly challenging in the case of replication where the doubled amount of DNA needs to be assembled into chromatin. It is of interest how cells achieve this assembly within biologically reasonable timescales.

A related physical process, the sequential adsorption of mutually exclusive particles from a bulk solution onto a lower-dimensional substrate, has been intensely studied in non-equilibrium statistical physics (9). In a simple one-dimensional model, sometimes referred to as the 'car parking' model, particles can bind to an initially empty line at any position where they do not overlap with particles already in place (10,11). If the adsorption is irreversible, all gaps larger than the particle size are quickly occupied and the coverage then runs into a 'jamming' plateau where nearly 75% of the line is covered (12). If the process is reversible, that is, if desorption is allowed, the density can be increased beyond this limit. Density increases then happen via rare events where a 'bad parker', a particle whose neighboring voids taken together are larger than the particle size, detaches and is replaced by two particles. This process is kinetically limited by the desorption rate, since at least one desorption event must precede any density increase. While the frequency of particles arriving at the substrate and attempting to adsorb must be much larger than the desorption rate to obtain high coverage, increasing it even further,

*To whom correspondence should be addressed. Tel: +49 (89) 289 - 12394; Fax: +49 (89) 289 -14656; Email: gerland@tum.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

e.g. via increase of the particle concentration in bulk solution, will not speed up the filling process. Instead, the adsorption rate merely sets the final density that is eventually achieved.

It was shown by Padinhateeri and Marko (13) that the jamming plateau can pose a serious kinetic challenge to the formation of dense nucleosome arrays: based on an *in vitro* measurement of the nucleosome formation rate (14) and a discrete version of the above-mentioned one-dimensional adsorption-desorption model that describes the nucleosomes as impenetrable particles covering 147 bp of DNA, they concluded that the physiological coverage of 90% of the DNA cannot be reached on biologically reasonable timescales without additional mechanisms. They also showed that an additional remodeling mechanism, which moves a nucleosome along the DNA in a randomly selected direction until it collides with its neighbor, can eliminate the kinetic problem and yield high coverage beyond the jamming plateau on much shorter timescales.

Here we show that the jamming problem is alleviated by the 'softness' of nucleosomes and by replication-guided nucleosome packing. We consider nucleosomes soft when the full-size footprints of neighboring nucleosomes can overlap. Such overlaps can arise by two different means: first, nucleosomes are known to breathe, i.e. nucleosomal DNA partially unwraps from the histone core, leading to a dynamic footprint on the DNA. Thermal fluctuations are sufficient to mediate transient unwrapping *in vitro* (15–17), while adenosine triphosphate-dependent chromatin remodeling enzymes (18) also affect unwrapping *in vivo* (19–21). Second, nucleosome assembly occurs in a stepwise manner, with an H3/H4 tetramer deposited first, followed by the addition of two H2A/H2B dimers (22). Effectively, the assembly process therefore leads to a transiently reduced DNA footprint. Taken together, nucleosome breathing and stepwise assembly permit neighboring nucleosome dyads to be more closely spaced than the canonical 147 bp footprint length, albeit with a reduced probability. There is indeed considerable genomic evidence for this behavior, including a direct experimental confirmation of the mutual invasion of neighboring nucleosomal DNA territories (23), and statistical evidence from the analysis of nucleosome maps (24,25).

Nucleosome softness necessitates a theoretical description that goes beyond treating nucleosomes as impenetrable 'hard-core' particles. A 'Soft-core Nucleosome Gas' (SoNG) model was previously introduced as a generalization of the Kornberg–Stryer model (26) for the analysis of gene-averaged steady-state nucleosome positioning patterns (24). To study the effects of nucleosome softness on the above-mentioned jamming problem, we introduce here a kinetic model that reproduces the steady state pattern of the SoNG model, but also describes the 'nucleosome filling dynamics', i.e. the approach to the steady state. We contrast the dynamic behavior of our kinetic SoNG model with that of the corresponding hard-core model that was previously studied (13,27). The comparison is justified, in the relevant parameter regime, by the observation that the steady-state nucleosome patterns of the two models are essentially the same and are both compatible with yeast data (24). We find that dense nucleosome arrays form much faster within the SoNG model, which proceeds via a new 'cramming' regime, avoiding the jamming behavior of hard-core nucleosomes.

In the second part of this article, we then explicitly consider nucleosome filling in the context of DNA replication. In yeast, replication starts at many origins across the genome and proceeds at rates that are highly variable, with 50 bp/s marking a typical speed of the replication fork (28). Most of the existing parental nucleosomes are directly passed on to the daughter strands (29). The essential steps of this 'segregation' process are removal and partial disassembly ahead of the fork, allocation to one of the daughter strands, and reassembly onto the daughter strand (30–32). *De novo* assembly of the missing nucleosomes then re-establishes densely packed chromatin behind the fork. The parental nucleosomes are deposited closely behind the fork (33) and typically within a distance of <400 bp from their parental loci (34), mediated by spatial association of the involved chaperones with the replication fork (31). Nucleosome positioning patterns appear to be virtually identical ahead of and behind the fork (35). There is some evidence that nucleosomes are distributed between the daughter strands in a random fashion with roughly equal shares (33,36,37), but the details of this process are still unclear and appear to be context dependent (38). Also, the process may differ between the leading and the lagging strand (39).

Taken together, the experimental evidence suggests that the segregation of parental nucleosomes is an orchestrated process that happens in close proximity to the fork, while the *de novo* deposition is less coordinated and less spatially constrained. We explore the interplay between a processively moving replication fork, nucleosome segregation and *de novo* assembly of nucleosomes on the newly synthesized DNA. We devise two simplified model scenarios intended to expose generic consequences of this interplay. Our model illustrates that nucleosome filling is facilitated by replication guidance in two cases, (i) if nucleosome (re)assembly behind the replication fork is sufficiently rapid to suppress the transient occurrence of large gaps and (ii) if the segregation process is highly orchestrated, e.g. such that the placement of parental nucleosomes alternates between the daughter strands and their new positions are highly correlated with their parental positions. In the first case, jamming is avoided by sequential filling, while positional 'memory' circumvents jamming in the second case.

MATERIALS AND METHODS

We first introduce our model for the assembly of nucleosome arrays on a large segment of naked DNA, and then extend it to a minimal model of replication-guided nucleosome filling. The starting point for the construction of our model, and a point of reference for its analysis, is a previously studied assembly model for nucleosome arrays (13), which is a kinetic version of the Kornberg–Stryer model (26). The DNA is represented as a one-dimensional lattice, with each lattice site representing a single base pair. Nucleosomes are assembled and evicted at random locations with the only constraint that gaps <147 bp wide cannot be filled. The steady-state of these kinetics is the statistical nucleosome distribution of the Kornberg–Stryer model, with the steady-state nucleosome density $\overline{\rho}$ set by the rates for assem-

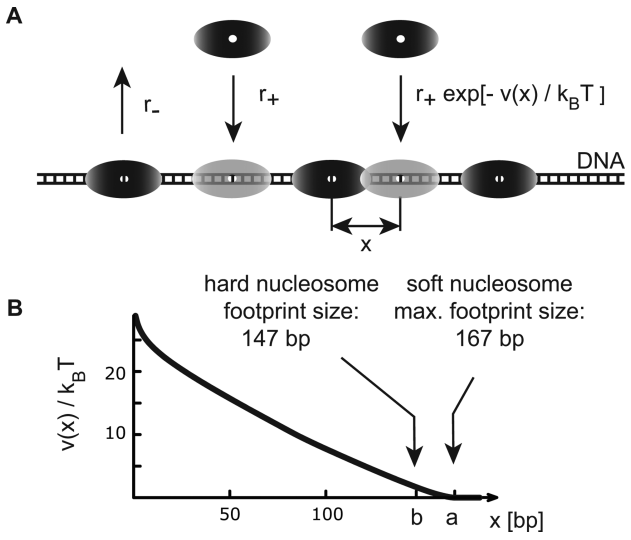


Figure 1. Kinetic model for the assembly of soft-core nucleosomes into dense arrays. (**A**) Nucleosomes (illustrated by shaded ellipses) are assembled onto free DNA at a rate r_+ and are evicted at a rate r_-. Nucleosomes can also form if this leads to an overlap of their DNA footprints, but the on-rate is then reduced by the Boltzmann factor of the interaction energy $v(x)$. Here, x denotes the distance between dyads, i.e. the center positions within nucleosome footprints (hollowed out in the nucleosome symbol). (**B**) The interaction potential $v(x)$ of soft-core nucleosomes has a range of $a = 167$ bp, as opposed to hard-core nucleosomes which obey strict exclusion over a range of $b = 147$ bp.

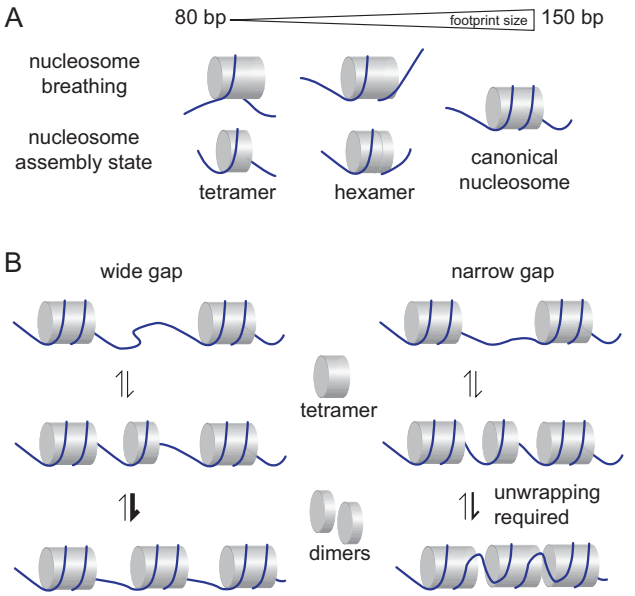


Figure 2. Nucleosome states with different DNA footprints and configuration-dependent nucleosome assembly. (**A**) Due to nucleosome breathing and multi-step assembly, nucleosomes have a multitude of internal states with a spectrum of DNA footprint sizes, i.e. they are 'soft'. In terms of DNA footprint size, the tetramer and hexamer states are equivalent to full nucleosomes with varying degree of DNA unwrapping. (**B**) The assembly rate of soft nucleosomes depends on the size of the gap between existing nucleosomes. Effectively, the stepwise assembly into a wide gap is faster than into a narrow gap, due to the required DNA unwrapping in the latter case.

bly and eviction, r_+ and r_-, via $\overline{\rho} = \eta/(1+\eta)$ with $\eta \exp(\eta) = r_+/r_-$, such that the steady-state density is a slowly increasing function of the rate ratio (11).

As illustrated in Figure 1A, we modify this model to take into account the softness of nucleosomes. The softness is manifested in a gradual repulsive potential between nucleosomes, shown in Figure 1B, which makes the assembly rate configuration-dependent, as opposed to the homogeneous assembly rate for hard-core repulsion. The rationale of our model is illustrated in Figure 2. Nucleosome softness stems from a multitude of structural states with different DNA footprint sizes, as depicted in Figure 2A. The statistical distribution of DNA footprint sizes is the essential determinant for the likelihood of assembling a nucleosome into a gap between existing nucleosomes. Within our model, the assembly into a narrow gap is less likely, but not excluded, as illustrated in Figure 2B.

The existence of nucleosome states with different DNA footprints is experimentally well established. Figure 2A depicts two classes of such states, those due to nucleosome breathing and those due to stepwise nucleosome assembly. Nucleosome breathing has long been suggested based on experiments showing that sites within the nucleosome footprint are accessible to binding proteins (15,40). Subsequently, the transient partial unwrapping of nucleosomal DNA from the histone core was directly demonstrated using single-molecule fluorescence techniques (16,17), and neighboring nucleosomes were shown to be capable of invading each others canonical 147 bp footprint (23). While these experiments were performed with reconstituted nucleosomes (via salt-gradient dialysis), indicating that thermal fluctuations are sufficient to produce transient unwrapping, ATP-dependent chromatin remodeling enzymes can also influence unwrapping *in vivo* (20,21). For our quantitative model, detailed below, it is important to note that nucleosome breathing leads to a rapid sampling of nucleosome states with different DNA footprints, with timescales in the millisecond to second regime (17).

According to the standard model for *in vivo* nucleosome assembly, an H3/H4 tetramer is first deposited onto the DNA and then completed to a full nucleosome by the addition of two H2A/H2B dimers (22,41). While there are different assembly pathways, both replication-coupled and replication-independent, involving different chaperones, remodelers and histone variants (22,30,31,42–44), the existence of assembly intermediates with reduced DNA footprint appears to be universal. Figure 2B (left panel) shows a simplified illustration of nucleosome assembly into a wide gap, where the neighboring nucleosomes do not impose any constraints on the assembly. The involved chaperones and remodelers are not shown in the illustration. In our quantitative model, we will use effective rates, which subsume all factors that affect these rates. Both assembly steps, formation of a tetrasome and integration of the heterodimers, are depicted as reversible processes; the backward reactions correspond to nucleosome eviction (18,45,46). Given that fully assembled nucleosomes are much more stable than assembly intermediates, the second step must be biased in the forward direction.

The right hand panel of Figure 2B illustrates the assembly into a narrow gap, where the neighboring nucleosomes affect the assembly process. This should primarily affect the rate of heterodimer integration in the second step, which requires DNA unwrapping in at least one of the shown nucleosomes (it does not matter how the required amount of unwrapping is distributed between the nucleosomes). The general implication for the construction of our quantitative model below is that soft nucleosomes, which feature an intrinsically variable DNA footprint size, will assemble at a rate that depends on the size of the gap between the existing neighboring nucleosomes, no matter if the newly deposited nucleosome enters as a breathing nucleosome or as some intermediate of stepwise assembly. For our purpose of modeling the process of nucleosome array formation, it is then adequate to coarse-grain the assembly of a single nucleosome into a one-step process with a configuration-dependent rate. This leads us to the kinetic model illustrated in Figure 1A, with a constant nucleosome eviction rate r_- and a basal assembly rate r_+ that is modulated depending on the distances to the neighboring nucleosomes (we also consider the effect of a variable eviction rate, see below). Starting from an initial configuration, we are then interested in the approach to 'equilibrium' of this model, which corresponds to an ATP-dependent non-equilibrium steady-state of the real system (47).

We incorporate the effect of the neighboring nucleosomes onto the nucleosome assembly rate via a repulsive 'soft-core' interaction potential $v\,(x)$, which gradually reduces the rate with decreasing distance x. This constitutes an adiabatic approximation, justified by the observation that nucleosome breathing samples nucleosome states with different DNA footprint sizes on a rapid timescale (17). The form of $v\,(x)$ should be chosen such that it (i) relaxes the widespread assumption of hard-core nucleosomes, (ii) leads to steady-state nucleosome distribution compatible with *in vivo* nucleosome patterns and (iii) produces a spectrum of DNA footprints compatible with experiments probing the accessibility of nucleosomal DNA. The task of finding such a potential was already carried out in (24). There, a soft-core potential with two free parameters, representing the maximal nucleosome footprint size and the effective stiffness of nucleosomes, was used to construct a statistical model, the 'SoNG', to describe *in vivo* nucleosome patterns. It was found that this model provides a more consistent description of the nucleosome patterns across multiple yeast species than the corresponding hard-core nucleosome gas. The fit to twelve different yeast species led to an effective interaction footprint of $a = 167$ bp and a nucleosome stiffness of $\varepsilon = 0.15 k_B T$ per base pair. The latter is consistent with estimates obtained from *in vitro* nucleosomal DNA accessibility data (48). The former is 20 bp longer than the DNA within a nucleosome core particle, which is not surprising, since steric constraints should disfavor neighboring nucleosomes with no linker DNA in between. Figure 1B shows the best-fit potential $v(x)$, with x measuring the dyad-to-dyad distance between neighboring nucleosomes.

The explicit shape of the potential is derived from the assumption of a constant energetic cost ε per bp to reduce the maximal DNA footprint a of a nucleosome at each end. For two neighboring nucleosomes placed with a dyad-to-dyad distance $x < a$, the required total footprint reduction can be distributed between the two nucleosomes and all possibilities are statistically weighted with the Boltzmann factor. Reference (24) finds that the simple expression

$$v(x) \approx (a - x)\epsilon - k_B T \ln\left[1 + (a - x)\left(1 - \mathrm{e}^{-\epsilon/k_B T}\right)\right] \quad (1)$$

for $x \leq a$ and $v(x) = 0$ for $x > a$ is an excellent approximation to this statistical average. Given that the values of its parameters a and ε were determined from *in vivo* data and yield a consistent description of gene-averaged nucleosome patterns over a range of yeast species, this potential essentially captures all effects that contribute to these steady-state patterns, including the action of remodelers. However, the fact that it is also consistent with the *in vitro* site-accessibility data suggests that thermal nucleosome breathing already leads to a similar sampling of footprint sizes as all *in vivo* processes combined. Therefore, equilibrium statistical models can be an adequate coarse-grained description of the 10-nm chromatin fiber even under *in vivo* conditions where it is an active system, contrary to what has been claimed elsewhere (27). We also note that our parameters for the maximal footprint and the unwrapping energy per base pair are compatible with those independently estimated from a different dataset (25). This analysis also considered an additional oscillatory component in the unwrapping energy to reflect the known 10–11 bp periodic preference of nucleosome positions. However, the additional oscillatory component has a minor effect on the statistics of nucleosome positions, visible only in the statistics of the relative rotational positions of neighboring nucleosomes (25), but not in the average nucleosome patterns considered here.

The nucleosome assembly model introduced here defines the kinetics for the SoNG model of (24). The soft-core potential modulates the local nucleosome assembly rate into the effective rate

$$\tilde{r}_+ = r_+ \, e^{-[v(x_\mathrm{L}) + v(x_\mathrm{R})]/k_B T}, \quad (2)$$

where x_L and x_R are the dyad-to-dyad distances to the next nucleosome on the left and right, respectively. As above, r_+ denotes the rate of nucleosome assembly at a given position sufficiently far away from existing nucleosomes. While we also simply refer to it as the 'on-rate', it should be noted that r_+ is not an association rate in the usual sense, since it does not measure the frequency of binding events per concentration of binding molecules but already includes this concentration. In fact, it is currently unclear (to the best of our knowledge) whether r_+ is limited by the number of free histones or by the number of chaperones that assist the formation of nucleosomes. Our expression for the effective rate $\tilde{r}_+$ assumes that the interactions with the left and right neighbor are additive, which requires the unwrapping of the two DNA ends of a nucleosome to be independent. Although correlations (or anti-correlations) are certainly possible, e.g. due to cooperative effects or electrostatic interactions, we currently do not know of any clear experimental evidence indicating any significant correlation of this kind.

To reproduce the steady-state of the SoNG model, the nucleosome eviction rate (or 'off-rate') must be constant, such that $\tilde{r}_+/r_-$ is proportional to the Boltzmann factor of the interaction potential. We have chosen to assign the entire Boltzmann factor to the on-rate in Equation 2, but we will

also consider other choices in the 'Results' section. There, we show that assigning the entire Boltzmann factor to the on-rate is the most conservative choice, in the sense that another choice would support our claim of substantially faster nucleosome filling of soft nucleosomes even more. Finally, given that we do not study sequence-guided nucleosome positioning at specific genes here, Equation 2 neglects any dependence of the on-rate on the DNA sequence. For hard-core nucleosomes, such effects have been considered in (13).

Model extension for replication-guided nucleosome filling

To model nucleosome filling in the wake of a moving replication fork, we consider the same kinetics as described above, but with a system size L (length of the DNA in bp) that increases with time. Specifically, we add lattice sites at the right hand boundary with a constant rate v_{repl}. This moving boundary represents the replication fork, and nucleosomes can only form behind the fork. The interaction between nucleosomes and the replication fork is described in the same way as the interaction between nucleosomes, i.e. the boundary condition is as if an imaginary nucleosome would always sit with its dyad at position $L + 1$ (with L increasing in time). The same boundary condition is applied at the fixed left boundary, i.e. there is an imaginary nucleosome with dyad at lattice site 0.

We first assume that parental and newly synthesized nucleosomes are mixed in a common pool, but later separate them in order to study how memory of parental nucleosome positions can affect the filling kinetics. To that end, we initialize our model in a partially filled state, which is constructed using a minimal model for the inheritance of nucleosome positions: we assume that during replication each daughter strand obtains half of the parental histones, on average. Starting from a parental nucleosome configuration according to the steady-state, each nucleosome is placed in its exact parental position on one of the daughter strands, proceeding along the parental DNA. If the previous nucleosome was placed on strand 1, the next one is place on strand 2 with the 'alternation probability' α, and *vice versa*, such that $\alpha = 0.5$ corresponds to a completely random placement, while $\alpha = 1$ corresponds to perfect alternation. We then use one of these daughter strands as the initial state for nucleosome filling.

Model implementation and parameter choice

To determine the nucleosome filling kinetics, we perform kinetic Monte Carlo simulations using the Gillespie algorithm (49). In these simulations, the state of the system is specified by the list of bound nucleosome dyad positions. The transition rates between states depend on the configuration as described above. For the DNA lattice, we either use a linear geometry with hypothetical fixed particles at the two ends (for replication-guided nucleosome filling, see above) or periodic boundary conditions corresponding to circular DNA (in all other cases). In each case, the total DNA length L is chosen large enough to ensure that none of our observables display a significant finite size effect. Except for the case of inherited parental nucleosomes (see above), we use an empty lattice as the initial condition. All observables are averaged over a sufficient number of simulation runs to extract the mean kinetics (see Supplementary Table S2 for details). As our main observable, we calculate the average nucleosome density $\rho(t)$ as the number of nucleosomes divided by the total DNA length, i.e. it can be interpreted as the inverse of the average spacing between nucleosomes (in bp). The quantity $\rho(t)$ also corresponds to the average probability for a base pair to be occupied by a nucleosome dyad. As a second observable, we calculate the time-dependent nucleosome pattern close to a boundary or a reference nucleosome.

Table 1. Soft-core nucleosomes require a different ratio of on- to off-rate than hard-core nucleosomes to obtain the same steady-state nucleosome density. At high densities, the required ratio is dramatically larger for hard-core nucleosomes.

Model	$1/\bar{\rho}$ [bp]	r_+/r_-
Hard-core nucleosomes	155	3.69×10^6
	165	149
	180	2.37
Soft-core nucleosomes	155	4.57×10^3
	165	200
	180	5.70

While the steady-state nucleosome density depends only on the rate ratio r_+/r_-, we also need one of these rates in absolute terms in order to estimate the timescale of the filling kinetics. From constant-force measurements of the shrinking rate of DNA during nucleosome assembly in *Xenopus* egg extract (without ATP), the on-rate r_+ was estimated at $\sim 12\, s^{-1}$ for a DNA segment with length of one nucleosome footprint (14). However, this can at most be an order of magnitude estimate for our purposes, since the rate may be significantly different *in vivo* and in the presence of ATP (50). In all figures of this manuscript, time is therefore measured in units such that $r_- = 1$, and only a rough correspondence to real time is indicated. To obtain different asymptotic nucleosome densities, we adjust the rate ratio r_+/r_- as listed in Table 1. Note that the hard-core model requires a dramatically higher r_+/r_- ratio at high densities than the soft-core model. This observation is consistent with the previous finding that the hard-core model does not provide a unified physical description of the 10-nm chromatin fiber in different yeast species (24).

RESULTS

We now use the quantitative model described above to address our two main questions: how does the softness of nucleosomes affect the kinetics of nucleosome assembly into dense arrays? And under which conditions can the replication process assist in the rapid reassembly of dense nucleosome arrays?

Kinetics of assembling soft nucleosomes into dense arrays

Figure 3 shows a characterization of the nucleosome filling kinetics for both the hard-core and the SoNG model (labeled by HaNG and SoNG, respectively). Panels A and B plot the average nucleosome density, $\rho(t)$, as a function of time (with a logarithmic time axis). For the time axis, we

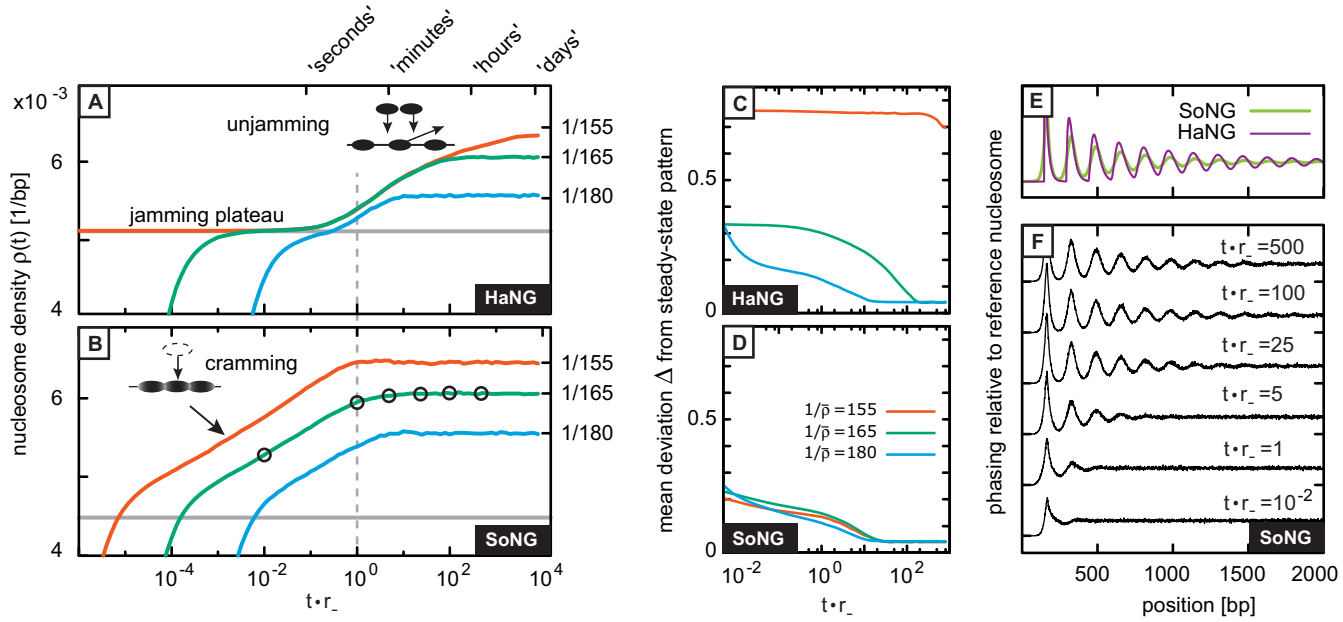


Figure 3. Characterization of nucleosome array assembly within the soft-core model and comparison to hard-core nucleosomes. (**A**) Time-dependent nucleosome density $\rho(t)$ for the HaNG model (with a logarithmic timescale). Three curves are shown for different final densities, as labeled on the right vertical axis. The gray horizontal line indicates the jamming density at which, statistically, non-overlapping adsorption opportunities have been exhausted. From this point, the hard-core nucleosomes must wait for desorption events to further increase the density (see sketch). The time labels above the figure mark approximate timescales based on the *in vitro* estimate of (13) (see main text). (**B**) Same plot for the SoNG model. Soft-core nucleosomes (represented by shaded ellipses in the sketch) have a larger maximal DNA footprint and accordingly have a smaller nominal jamming density (again indicated by a gray line). However, after reaching this density, $\rho(t)$ does not plateau, but instead enters a 'cramming' stage. (**C**) Mean deviation $\Delta(t)$ between the time-dependent nucleosome pattern $p(x, t)$ and its steady-state $\overline{p}(x)$ for the HaNG model (see text for details). The three curves correspond to the same cases as shown in panel (A). (**D**) Same plot for the SoNG model. (**E**) Comparison of the equilibrium patterns of the two models at a nucleosome density of 1/(165 bp). (**F**) Snapshots of the average nucleosome pattern during the filling process of the SoNG model for 1/(165 bp) final nucleosome density. The circles on the green line in panel B mark the points in time for which the snapshots are displayed.

use the average dwell time of a nucleosome on the DNA, $1/r_-$, as our time unit. The shape of the filling curves of the SoNG model in Figure 3B is strikingly different from that of the HaNG model in Figure 3A: initially, when the DNA coverage is still low, the density increases rapidly (linear in time) in both cases, since nucleosome–nucleosome interactions play a negligible role. Then, however, the filling of the hard-core model stalls, while the density steadily increases for the soft-core model. The stalling of the HaNG model occurs at a density of 1/197 bp, the 'jamming plateau' marked by the gray line. This plateau corresponds to a DNA coverage equal to the nontrivial theoretical limit of 74.8… % calculated by Rényi (12) for the irreversible random binding of equally sized objects to a continuous one-dimensional substrate. The continuum limit provides an accurate description also for our discrete DNA substrate, since nucleosomes are large compared to the discrete length unit of a single base. Importantly, the jamming plateau is independent of the on-rate: the three different curves in Figure 3A correspond to different r_+ values, but stall at the same coverage level, albeit at different times.

For the green traces in both Figure 3A and B, the on-rate is adjusted such that a steady-state nucleosome spacing of 165 bp is ultimately reached, corresponding approximately to the average spacing in *Saccharomyces cerevisiae* under physiological conditions (6). The blue traces illustrate the behavior for a reduced density with a 180 bp spacing, while the red traces illustrate the case of close to maximal packing with a 155 bp spacing. In the case of the HaNG model, nucleosome filling beyond the level of the jamming plateau starts at a timescale of $\sim 1/r_-$ for all three traces. This reflects the fact that 'unjamming' requires the removal of 'bad parkers', as illustrated in the sketch inside panel A. Note that for the green and red traces, the filling process is stalled for several orders of magnitude in time, and that unjamming is logarithmically slow (10).

Nucleosome filling within the SoNG model never stalls, but instead displays a 'cramming' stage. The crossover between the initial filling and the onset of cramming happens when most gaps are too small for further nucleosomes to attach at non-overlapping positions. This density is marked by the gray line in Figure 3B. The biophysics of the cramming process is analyzed in detail further below. On the phenomenological level, the cramming stage ends when the final steady-state density is reached, which happens on about the same timescale for all three traces. Remarkably, this timescale is not much longer than the unbinding timescale $1/r_-$, suggesting that breathing nucleosomes will reach dense packing already when hard-core nucleosomes are only starting the unjamming process. This conclusion is not dependent on the assumption, made in Equation 2, that the nucleosome–nucleosome interaction only affects the on-rate: as shown in Supplementary Figure S1, this assumption is conservative, since all other choices

for the dependence of the kinetics on the interaction (see Supplementary Text, Section I) only lead to even faster filling with soft-core nucleosomes.

Dynamics of nucleosome phasing

While the above analysis showed that a high density is reached very rapidly with soft-core nucleosomes, it did not address the question of whether the characteristic phasing of nucleosome arrays (7) forms on the same rapid timescale. We calculate the dynamics of nucleosome phasing via the time-dependent probability $p(x, t)$ of finding a nucleosome dyad at position x at time t, given that one reference nucleosome is fixed at position 0. To quantify the timescale of the approach to steady-state, we compute the normalized mean deviation Δ between the time-dependent pattern $p(x, t)$ and its steady-state limit $\overline{p}(x)$,

$$\Delta(t) = \frac{1}{L\overline{\rho}} \sum_{x=1}^{L} |p(x, t) - \overline{p}(x)| \; . \qquad (3)$$

Here, the steady-state limit $\overline{p}(x)$ is calculated exactly by the transfer matrix method as previously described (24), while $p(x, t)$ is obtained as an average over many kinetic simulations, see 'Materials and Methods' section and Supplementary Table S2.

Figure 3C and D show how the mean deviation Δ decreases as a function of time, for the HaNG and the SoNG model, respectively. The three different curves in each panel correspond to the same $\overline{\rho}$ values considered already in panels A and B. These curves demonstrate that the dynamics of the phasing pattern largely follow the behavior of the nucleosome density $\rho(t)$. For the HaNG model, the timescale of the approach to steady state increases dramatically with the steady-state density $\overline{\rho}$, whereas it displays no significant dependence on $\overline{\rho}$ for the SoNG model. Figure 3E superimposes the final pattern $\overline{p}(x)$ for the HaNG and SoNG model to show that the steady states of the two models are compatible with each other. Note that the sharp peaks in the pattern of the HaNG model are due to the exact positioning of the reference nucleosome; after taking into account the fuzziness in the positioning of e.g. the +1 nucleosome, the two patterns become almost identical and are both compatible with the genome-averaged experimental pattern of *S. cerevisiae* (24).

Figure 3F shows several snapshots of the dynamics of nucleosome phasing within the SoNG model, corresponding to the points marked by circles on the green curve in panel B. These snapshots illustrate how the phasing pattern emerges on the unbinding timescale $1/r_-$, by gradual propagation of the pattern from the reference nucleosome. Supplementary Figure S2 shows similar plots for other densities, and also for the HaNG model, which displays much slower dynamics consistent with Figure 3C. The slow dynamics of nucleosome pattern formation within the HaNG model were also observed in a recent study and motivated the extension of the model by an additional remodeling mechanism (27).

Physical analysis of the cramming stage

During the cramming stage in the nucleosome filling dynamics of Figure 3B, nucleosomes are 'squeezed' into progressively shorter gaps. The cramming stage begins when the average density $\rho(t)$ reaches the jamming density marked by the gray line in Figure 3B, i.e. the jamming density $\rho_c = 0.748/a$ for the maximal footprint a of the soft-core nucleosomes. At this point, the density $G_c(x)$ of inter-nucleosome gaps of size x has the shape shown in Figure 4A. Here, x is measured as the distance between the dyads of neighboring nucleosomes and the gap density is plotted on a logarithmic axis. Note that despite the overall dependence of the filling dynamics on r_+, the distribution $G_c(x)$ at the onset of cramming is invariant, as shown in Supplementary Figure S3. To a good approximation, the shape of $G_c(x)$ is piecewise exponential with three regimes (see also Supple-

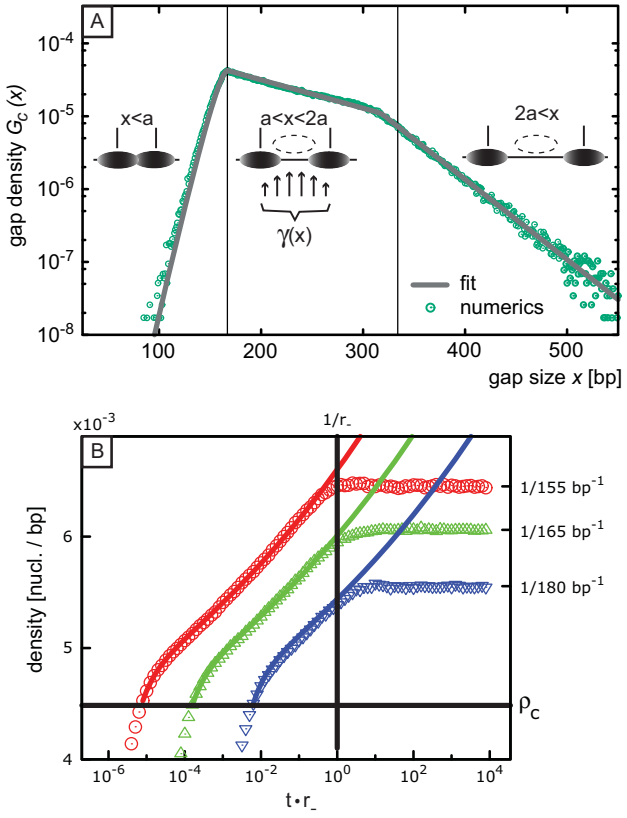


Figure 4. Analysis of nucleosome filling during the cramming stage of the SoNG model. (**A**) Length-dependent gap density, $G_c(x)$, at the onset of cramming on a logarithmic axis (the gap length x is measured as the distance between the dyads of neighboring nucleosomes). The green symbols show simulation data whereas the gray line shows the fit described in the main text. Vertical lines separate three regimes (see sketches for illustration): for $x < a$, neighboring particles already overlap, such that intervening adsorption will be strongly suppressed, due to an enormous energy penalty. The second regime, $a < x \lesssim 2a$, largely determines the cramming dynamics, with gaps that provide less than a nucleosome footprint of free DNA space. In the third region, $x \gtrsim 2a$, gaps are large enough to permit the assembly of an intervening nucleosome without interaction; these gaps will fill very quickly. All gaps fill with rate $\gamma(x)$ defined in Equation 5; the middle sketch illustrates how the different attachment possibilities within a gap sum up to the gap's total filling rate $\gamma(x)$. (**B**) Nucleosome filling during the cramming stage. Equation 4 (lines) describes the Monte Carlo simulations (symbols) throughout the cramming stage. See Table 1 for parameters. The horizontal gray line indicates the cramming density ρ_c, while the vertical gray line indicates the desorption timescale $1/r_-$, which marks the end of the cramming stage.

mentary Text, Section II): (i) for $x < a$, where the neighboring nucleosomes already overlap, the gap density increases roughly exponentially with gap size. The slope (gray line in Figure 4A) follows directly from the Boltzmann factor for the interaction potential. The addition of nucleosomes into gaps in this regime is exceedingly unlikely. (ii) For $x \geq a$, but less than an upper threshold $x^* \approx 2a$, the neighboring nucleosomes do not overlap, but the gap does not allow a new nucleosome to bind without significant interaction with at the least one of the existing nucleosomes. In this regime, $G_c(x)$ decays slowly with increasing gap size. (iii) For $x > x^*$, the gaps are wide enough to fit a new nucleosome without significant hindrance, such that only very few of these gaps remain at the onset of the cramming stage. Accordingly, the distribution $G_c(x)$ decays rapidly with x in this regime.

If the dominant kinetic process during the cramming stage is the filling of initially created gaps, we should be able to predict the dynamics of $\rho(t)$ from $G_c(x)$. Specifically, the time-dependent nucleosome density should follow

$$\rho(t) = \rho_c + \sum_x G_c(x) \left(1 - e^{-\gamma(x)(t-t_c)}\right) , \qquad (4)$$

where t_c is the time at which the nucleosome density reaches the cramming threshold ρ_c (defined above). As the total number of gaps in the system equals the number of nucleosomes, we also have the relation $\sum_x G_c(x) = \rho_c$. The factor in brackets corresponds to the probability that a gap of size x has been filled at time t, given a gap filling rate $\gamma(x)$. This rate can be obtained as the sum over the attachment rates at all positions between the dyads, with each position-specific rate according to Equation 2, such that

$$\gamma(x) = r_+ \sum_{x'=1}^{x-1} e^{[-v(x')-v(x-x')]/k_BT} \qquad (5)$$

(see also the sketch inside regime (ii) of Figure 4A). The cramming dynamics predicted by Equation 4 are shown in Figure 4B alongside the three simulated filling curves, displaying excellent agreement until the characteristic timescale $1/r_-$ for unbinding when the cramming stage ends. For the numerical evaluation of Equation 4, we used the entire distribution $G_c(x)$ (described as the piecewise exponential shown as the gray line in Figure 4A; see Supplementary Text, Section II for more details), however regime (ii) of the gap size distribution governs the cramming behavior (see Supplementary Figure S4 for an analysis of the relative contributions of the three regimes).

The above quantitative analysis confirms our biophysical interpretation of the cramming stage as a progressive filling of the gap distribution that is established prior to the onset of cramming. In other words, higher-order processes involving the recursive filling of newly created gaps are negligible. At the end of the cramming stage, when unbinding becomes relevant, the subsequent final equilibration occurs very fast in the SoNG model. This is due to the fact that after unbinding, the rebinding position of a nucleosome is not uniformly distributed within gaps, as is the case in the HaNG model. Rather, the energetic gradient of their interactions with neighboring nucleosomes provides a 'guiding funnel' toward proper spacing.

Nucleosome filling behind a moving replication fork

We now consider effects on the kinetics of nucleosome filling that can result from the process of DNA replication, using the model extension described in 'Materials and Methods' section. We first investigate a scenario where the DNA is cleared from nucleosomes by the replication fork and nucleosome filling of the daughter strands occurs behind the moving fork. The replication fork moves along the DNA with a certain speed v_{repl} and thereby exposes newly synthesized DNA continuously. Once the synthesized DNA has reached a sufficient length, a new nucleosome can assemble. If the progression speed is slow and the adsorption rate large, this will occur almost immediately when the new strand is long enough. Replication then continues and soon the next nucleosome can attach adjacent to the previous one. This leads to closely packed nucleosomes in the wake of the moving fork, as depicted in the upper part of Figure 5A. We refer to this as 'replication-guided' assembly. If, on the other hand, the replication fork progresses rapidly compared to the assembly rate, then large stretches of newly synthesized DNA will be exposed to nucleosome assembly at random positions, leading to jamming, as sketched in the lower part of Figure 5A. Note that in the limit of very fast replication, we would recover the replication-independent nucleosome filling studied above.

A typical replication speed is $v_{\mathrm{repl}} = 50$ bp/s or one-third of a nucleosome footprint per second, however a variation of at least a factor 10 in the replication speed has been observed (28). For the rate of nucleosome assembly, the extrapolated *in vitro* rate from (14) suggests 12 assembly attempts per second within a nucleosome footprint. Clearly, the *in vivo* rate could be substantially different and additionally modified by histone level and chaperone activity regulation. Given this spread in the relevant quantities, both of the above scenarios could be realistic. We therefore show a quantitative analysis of both regimes within the SoNG model in Figure 5.

Figure 5B shows the nucleosome density in the wake of the moving fork, while Figure 5C shows the time evolution of the density at a fixed position. In each case, a simulation result for a slow and fast replication fork is shown (see Supplementary Materials for more details). For slow replication (red traces), nucleosomes are packed tightly behind the fork. The density rises quickly after the fork has passed by and initially exceeds its ultimate equilibrium value. Later, the density decreases through events in which two nucleosomes leave and the gap is filled by only one nucleosome. In contrast, for fast replication (blue traces), the moving fork guides nucleosomes only weakly and the density follows the cramming behavior of naked DNA (black traces) in its last stage.

From the quantitative analysis in Figure 5 it becomes clear that the functionally ideal regime corresponds to an intermediate case where the replication-guided assembly directly leads to the steady-state density: if the assembly rate is tuned with respect to the replication speed, the nucleosomes behind the fork will already display the proper spacing, and the filling curve in Figure 5C will rapidly rise to the steady-state nucleosome density, without any overshoot or relaxation behavior. Such a fine-tuned optimal behavior

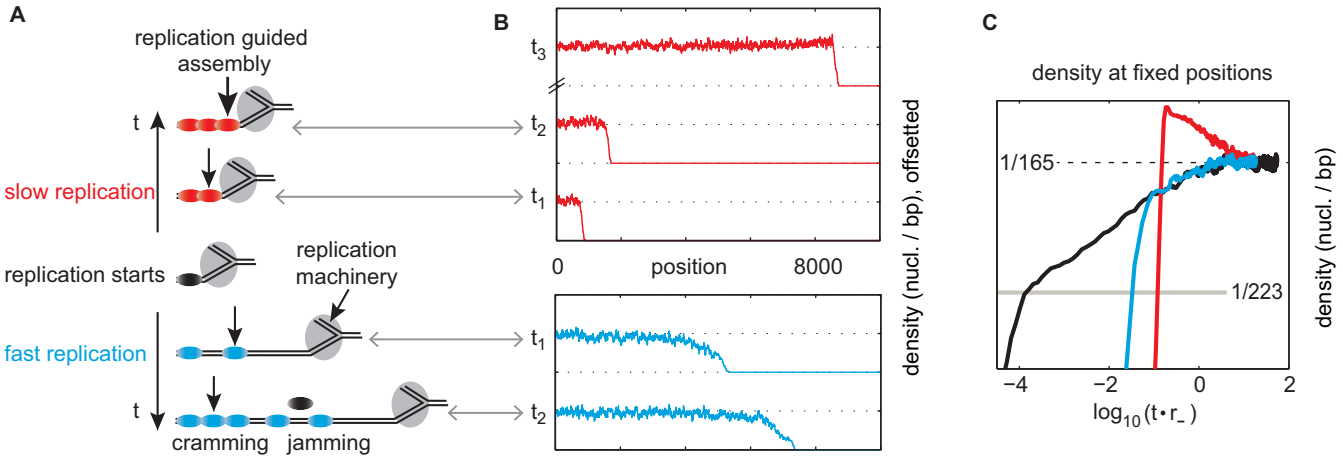


Figure 5. Illustration of the wake filling mechanism. (**A**) Schematic of different filling regimes in the wake of the moving replication fork (only one daughter strand shown). If the fork progresses slowly, or if replication machinery tightly replaces nucleosomes in its wake, nucleosomes attach to synthesized DNA very quickly once a sufficiently wide segment becomes available, leading to tight packing in the wake. If the fork progresses quickly, however, the newly synthesized DNA is left essentially empty. Jamming can then occur, which, for soft nucleosomes, is resolved by cramming. (**B**) Snapshots of the nucleosome density along the DNA at different times $t_1 < t_2 \ll t_3$. (**C**) Time-evolution of the nucleosome density at fixed positions. Time is set to zero when the fork passes by and filling can start. Red: for slow replication, the density initially exceeds and then approaches its equilibrium value. Blue: for fast replication the density in its final phase follows the replication-independent cramming behavior (shown in black for comparison). The cramming density ρ_c is indicated by the gray line. See Supplementary Figure S5 for detailed simulation parameters.

could be obtained, for instance, by coupling the replication speed to the concentration of free histones (see discussion for possible evidence).

Nucleosome filling guided by parental nucleosome positions

In the above analysis, parental and new nucleosomes were lumped together in one pool, from which nucleosomes were randomly placed on the newly synthesized DNA. To explore possible effects of nucleosome segregation on the filling kinetics, we now consider a variant of our model which takes the deposition of parental nucleosomes on the daughter strands into account. We assume that all parental nucleosomes are distributed between the daughter strands by the replication machinery, thus generating an initial state for the subsequent filling with new nucleosomes that is significantly different from the empty DNA assumed above. Given that important aspects of the nucleosome segregation process are not yet experimentally characterized, we consider only extreme scenarios that illustrate the potential effects most clearly. Specifically, we consider only the fast replication regime of the previous section, which allows for a clear separation between nucleosome segregation and DNA filling with new nucleosomes (we do not need to consider the moving replication fork explicitly, but can use the deposited parental nucleosomes as an initial condition for the filling with *de novo* assembled nucleosomes). Furthermore, we assume the idealized case where a segregated nucleosome receives the same position on a daughter strand as it had on the parental DNA.

We focus on the question of how the filling kinetics are affected by the splitting process that distributes the parental nucleosomes between the two daughter strands, see 'Materials and Methods' section. Figure 6A depicts the two extreme cases of random distribution (bottom) and perfectly alternating distribution (top). Figure 6B and C show the corresponding filling kinetics, for the SoNG and the HaNG model, respectively. Figure 6B and C also show the filling curve for empty DNA as a reference (black lines). For the alternating initial distribution (red lines in Figure 6B and C) we observe that the density reaches its final value very quickly, avoiding jamming (for the HaNG model) and cramming (for the SoNG model). This is to be expected, given that all gaps have the correct size for one additional nucleosome. The gaps are quickly filled, re-establishing the equilibrium density with no need for rearrangements. For random allocation between strands (blue lines) there are gaps that have accommodated two or more nucleosomes on the parental DNA, which are then likely to become obstructed by 'bad parkers', thus leading to jamming. The effects of parental histone positions are more pronounced for the HaNG model where jamming is more severe than in the SoNG model. However, even for the SoNG model the final approach to the steady-state is considerably faster when nucleosome filling is guided by alternating nucleosome segregation.

DISCUSSION

Summary

Our model-based analysis has shown that the formation of dense nucleosome arrays is expedited by nucleosome softness, which can be attributed to nucleosome breathing and stepwise assembly, and by positional guidance obtained through the replication process. For the latter, we identified two different mechanisms, (i) positional guidance by a moving replication fork and (ii) positional guidance by segregated parental histones. Mechanism (i) provides optimal guidance, if the nucleosome assembly rate is matched to the speed of the replication fork. If, however, the replication machinery moves too fast to allow

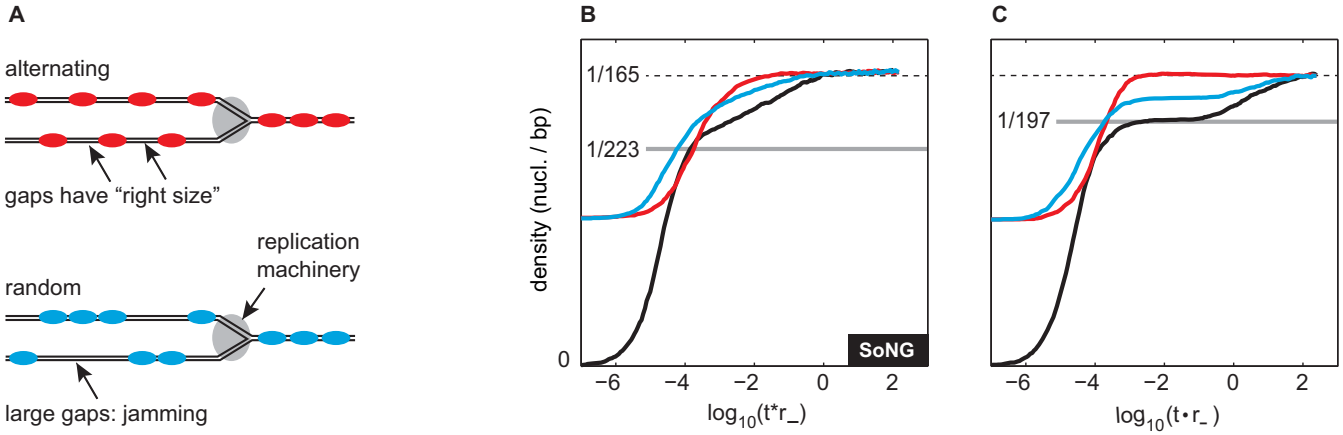


Figure 6. Effects of inherited parental nucleosome positions on filling dynamics. (**A**) An alternating distribution of parental nucleosomes onto the daughter strands ($\alpha = 1$) leads to gaps that each accommodate a single additional nucleosome (top panel), while a random distribution ($\alpha = 0.5$) leads to gaps of various sizes such that jamming will occur for *de novo* deposited nucleosomes (bottom panel). (**B**) Nucleosome filling dynamics of the SoNG model starting from an alternating inherited configuration (red), a random inherited configuration (blue) and from empty DNA as a reference (black). The dashed line marks the asymptotic density and the gray line the jamming density. (**C**) Same plot for the HaNG model.

dense packing in its wake, nucleosomes form at random positions, resulting in 'jammed' nucleosome configurations which must be resolved to further increase the DNA coverage. In this fast replication regime, mechanism (ii) can still prevent jamming. A caveat is that it is effective only if the segregated nucleosomes (approximately) retain their positions during DNA replication and neighboring nucleosomes on the parental DNA are typically segregated to different daughter strands. However, even if jammed nucleosome configurations do occur, we found that their effect on the nucleosome filling kinetics is much less dramatic for soft nucleosomes than when nucleosomes are approximated as hard-core particles.

We find it useful to illustrate replication-guided nucleosome packing by extending the 'car parking' analogy that is often used for one-dimensional adsorption-desorption models (11). Imagine a truck that slowly moves along the curb of a street, e.g. performing roadwork. Cars in search of parking spots can park right behind the truck as soon as the distance from the previously parked car is large enough. This results in densely spaced cars along the curb, analogous to the replication-guided filling described above. Now, assume that the truck moves by more than one car's length before a driver looking for a parking spot arrives. Then, parking will no longer be ordered behind the truck. If however, the roadwork performed by the truck consists of painting parking spot guidelines on the road, dense packing is again established (assuming equal car lengths and drivers respecting guidelines). In the nucleosome context, such positioning guides emerge if the gaps between inherited nucleosomes correspond directly to the space vacated by a parental histone, such that each gap can be quickly filled by a single assembly event.

That nucleosomes with an effective soft-core interaction display much faster filling kinetics than hard-core nucleosomes is not a trivial effect. They are larger than their counterparts in the HaNG model and the interaction parameters are determined such that the same equilibrium density and compatible nucleosome patterns are obtained in both cases (24). Furthermore, the filling kinetics of the soft-core nucleosomes differs significantly from that of the hard-core nucleosomes: the latter quickly run into a stagnating nucleosome density, the 'jamming plateau', followed by a long period of collective rearrangements during which the nucleosome density creeps to the final steady-state density. In contrast, the density of soft-core nucleosomes does not plateau before reaching the final steady-state. Rather, the SoNG model displays a cramming stage, during which the nucleosome density steadily increases, followed by a rapid relaxation to the steady-state nucleosome pattern via nucleosome rearrangements that are guided by the nucleosome–nucleosome interaction. Taking nucleosome softness into account thus leads to profound effects that should not be ignored in kinetic studies of nucleosome array formation. For steady-state properties, the hard-core description (26) remains a useful abstraction that can even quantitatively describe the gene-averaged nucleosome pattern adjacent to nucleosome-free regions (51), albeit not in a unified way, across different species (24).

Assumptions and limitations

Taken together, our findings suggest that the biophysics of nucleosomes and DNA replication helps cells to avoid the kinetic problem of jamming, which would otherwise arise in the formation of dense nucleosome arrays (13). We note that the dramatic speedup of the filling kinetics of the soft-core model compared to the hard core model is a *generic* property of soft nucleosomes, i.e. it is robust to changes in the specific shape of the interaction potential. Clearly, our coarse-grained model has many simplifying assumptions, which could potentially affect this conclusion. For instance, our SoNG model is restricted to nearest-neighbor interactions between nucleosomes. While it is certainly possible to integrate longer-range interactions (mediated by higher-order chromatin structure) into these models (52), there is currently no experimental evidence that nearest-neighbor

interactions do not suffice to describe the statistical distribution of nucleosome positions. Another simplifying assumptions was to ignore the DNA sequence preferences of nucleosomes, which would make our on- and off-rates r_+ and r_- position-dependent. However, position-dependent rates can provide additional guidance and further reduce the tendency of nucleosomes to jam, i.e. the assumption is not critical.

One of the limitations of our study is the inability of our coarse-grained model to predict the absolute timescale of the nucleosome filling kinetics. A crucial unknown parameter is the *in vivo* rate of nucleosome assembly (the rate r_+ in our model). Using the *in vitro* estimate of (14), our SoNG model suggests that 90% DNA coverage can be reached on a timescale of several minutes, see Figure 3. However, this estimate has considerable uncertainty, due to both the extrapolation to zero applied force in (14) and the usage of an *in vitro* rate for an *in vivo* process. The strength of the coarse-grained modeling approach is that it can compare different mechanisms in terms of their relative speed of nucleosome array assembly. Our model comparison demonstrated that soft-core nucleosomes have a kinetic advantage over hard-core nucleosomes, with an almost two orders of magnitude shorter assembly time for nucleosome arrays with the spacing of *S. cerevisiae*. The SoNG assembly time is in fact comparable to the timescale estimated within a similar coarse-grained model for hard-core nucleosomes, but with remodeler-assisted nucleosome sliding (13) (passive nucleosome sliding without assistance by remodeling enzymes is too slow to significantly affect assembly times).

Our treatment of replication-guided nucleosome packing has introduced a minimal model for the simultaneous kinetics of DNA replication and assembly of nucleosome arrays.

Rather than modeling the detailed processes at the replication fork, we focused on the more general question of how the processive nature of replication influences filling kinetics. This establishes the basis upon which more elaborate orchestration of chromatin reassembly may operate. Slow progression of the replication fork compared to the nucleosome assembly rate helps to avoid nucleosome jamming, while the guiding effect of the fork is negligible at fast progression speeds. We showed that for very slow progression, the density in the wake of the fork even exceeds its equilibrium value temporarily. It is tempting to speculate whether a coupling mechanism between the assembly rate and the fork speed might allow cells to tune the replication-guided density to the steady-state value. This density would then be reached substantially faster than even the soft-core model predicts for the homogeneous case. An indication for such a feedback mechanism was indeed reported in mammalian cells: limiting the supply of new histones slows down the replication fork (53). However, in these experiments the nucleosome density on the nascent daughter strands was reduced in histone depleted conditions, indicating that the feedback mechanism might not fully compensate for the lower histone availability. To what extent this feedback might tune the ratio of speed and assembly toward the optimal regime is currently unclear.

We also investigated possible influences of histone segregation. Again, our model is certainly not meant as a full description of these processes in yeast cells. For instance, we have ignored differences between the replication of the leading and the lagging strand. Instead, our model serves to illustrate generic consequences of an interplay between the DNA replication machinery and nucleosome assembly. This interplay depends on the statistical properties of the nucleosome segregation process. Fast assembly of dense arrays is facilitated by an alternating deposition of parental nucleosomes to the daughter strands and a high correlation between parental nucleosome positions and those on the daughter strands. These aspects of DNA replication and nucleosome segregation are insufficiently characterized experimentally. However, it does seem clear already that the details are context-dependent. For instance, while nucleosomes are generally believed to be allocated in equal shares to the two daughter strands, nucleosomes in *Drosophila* germline stem cells are mainly segregated to one daughter strand, while *de novo* assembled nucleosomes are enriched on the other (38). It is also interesting to note that replication of the lagging strand was found to be tied to the assembly of nascent chromatin: Okazaki fragment lengths are multiples of the average nucleosome spacing in *S. cerevisiae* with fragments terminating preferentially at consensus dyad positions, and suppression of nucleosome assembly resulted in longer fragments (39).

Finally, our model does not explicitly account for remodeling enzymes, which are known to reposition, remove and restructure nucleosomes (18). As described in the model section, we incorporate the action of histone chaperones into our effective assembly rate r_+, and the ATP-assisted removal of nucleosomes in our effective eviction rate r_-. Note that the parameters of the potential which describes the reduced on-rate for overlapping adsorptions are determined from *in vivo* nucleosome positioning patterns that include the effects of remodelers. Our analysis has shown that reasonable effective rate constants r_+ and r_- are sufficient for rapid formation of dense nucleosome arrays, once nucleosome softness is taken into account. While our findings don't deny that other types of remodeling processes are also taking place, they suggest that active lateral repositioning of nucleosomes is not required to form dense nucleosome arrays in a timely manner. This conclusion is not in conflict with the observation that the reconstitution of nucleosome patterns across the 5' ends of yeast genes requires whole cell extract and ATP (54), given that eviction of nucleosomes is also ATP-assisted.

However, these and other experiments at reduced nucleosome density show that our model is not sufficient for a coherent quantitative description of gene-averaged nucleosome patterns for all experimental conditions. At a minimum, this will require either a mechanism that pushes nucleosome toward the 5′ ends of genes (54) or a mechanism that mediates a nucleosome–nucleosome attraction (24).

Experimental ramifications and outlook

Our theoretical analysis of replication-guided nucleosome packing stimulates several experimental questions. Given that the kinetically optimal scenario lies in the regime where the replication speed is tuned to the nucleosome assembly rate r_+ times the dyad-to-dyad spacing of the packed nucleosome array, it is of particular interest to test whether yeast

cells typically operate in this regime. The replication fork speed could even be controlled externally via hydroxyurea (55,35) to study different regimes of the wake filling mechanism *in vivo*. A useful experimental observable for comparison with theoretical models would be the nucleosome density profile as a function of the distance to the replication fork. This density profile should display a depletion zone behind the fork, whose width is a function of the ratio of the replication fork speed to the speed of nucleosome assembly.

The process of nucleosome segregation has been of great interest in the context of epigenetic inheritance of histone modifications. Our theoretical analysis has shown that the statistical properties of this process can also have strong effects during the process of reforming dense nucleosome arrays after DNA replication. An experimental analysis of these statistical properties could simultaneously shed new light on both questions and would be highly desirable. We hope that a dynamical analysis of nucleosome arrays will significantly advance our quantitative understanding of the processes that shape the arrays.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online.

FUNDING

Bavarian Research Network for Molecular Biosystems (BioSysNet); Nanosystems Initiative Munich (NIM); Funding for open access charge: Ulrich Gerland.
Conflict of interest statement. None declared.

REFERENCES

1. Venkatesh,S., Workman,J.L. and Smolle,M. (2013) UpSETing chromatin during non-coding RNA production. *Epigenetics Chromatin*, **6**, 16.
2. Svaren,J. and Hoerz,W. (1997) Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends Biochem. Sci.*, **22**, 93–97.
3. Bai,L. and Morozov,A.V. (2010) Gene regulation by nucleosome positioning. *Trends Genet.*, **26**, 476–483.
4. Bell,O., Tiwari,V.K., Thomä,N.H. and Schübeler,D. (2011) Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.*, **12**, 554–564.
5. Rando,O. and Winston,F. (2012) Chromatin and transcription in yeast. *Genetics*, **190**, 351–387.
6. Jiang,C. and Pugh,B.F. (2009) A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biol.*, **10**, R109.
7. Yuan,G., Liu,Y., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, **309**, 626–630.
8. Jiang,C. and Pugh,B.F. (2009)(2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.
9. Evans,J.W. (1993) Random and cooperative sequential adsorption. *Rev. Mod. Phys.*, **65**, 1281–1329.
10. Jin,X., Tarjus,G. and Talbot,J. (1994) An adsorption-desorption process on a line: kinetics of the approach to closest packing. *J. Phys. A*, **27**, L195–L200.
11. Krapivsky,P.L. and Ben-Naim,E. (1994) Collective properties of adsorption–desorption processes. *J. Chem. Phys.*, **100**, 6778–6782.
12. Rényi,A. (1963) On a one-dimensional problem concerning random space filling. *Sel. Trans. Math. Stat. Prob.*, **4**, 203–218.
13. Padinhateeri,R. and Marko,J.F. (2011) Nucleosome positioning in a model of active chromatin remodeling enzymes. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7799–7803.
14. Padinhateeri,R., Yan,J. and Marko,J.F. (2007) Nucleosome hopping and sliding kinetics determinedfrom dynamics of single chromatin fibers in Xenopus egg extracts. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 13649–13654.
15. Polach,K.J. and Widom,J. (1995) Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.*, **254**, 130–149.
16. Li,G. and Widom,J. (2004) Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.*, **11**, 763–769.
17. Buning,R. and van Noort,J. (2010) Single-pair FRET experiments on nucleosome conformational dynamics. *Biochimie*, **92**, 1729–1740.
18. Clapier,C.R. and Cairns,B.R. (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.*, **78**, 273–304.
19. Flaus,A. and Owen-Hughes,T. (2004) Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? *Curr. Opin. Genet. Dev.*, **14**, 165–173.
20. Fan,H., He,X., Kingston,R.E. and Narlikar,G.J. (2003) Distinct strategies to make nucleosomal DNA accessible. *Mol. Cell*, **11**, 1311–1322.
21. Kassabov,S.R., Zhang,B., Persinger,J. and Bartholomew,B. (2003) SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol. Cell*, **11**, 391–403.
22. Akey,C.W. and Luger,K. (2003) Histone chaperones and nucleosome assembly. *Curr. Opin. Struct. Biol.*, **13**, 6–14.
23. Engeholm,M., de Jager,M., Flaus,A., Brenk,R., van Noort,J. and Owen-Hughes,T. (2009) Nucleosomes can invade DNA territories occupied by their neighbors. *Nat. Struct. Mol. Biol.*, **16**, 151–158.
24. Moebius,W., Osberg,B., Tsankov,A.M., Rando,O.J. and Gerland,U. (2013) Toward a unified physical model of nucleosome patterns flanking transcription start sites. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5719–5724.
25. Chereji,R.V. and Morozov,A.V. (2014) Ubiquitous nucleosome crowding in the yeast genome. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 5236–5241.
26. Kornberg,R.D. and Stryer,L. (1988) Statistical distribution of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.*, **16**, 6677–6690.
27. Parmar,J.J., Marko,J.F. and Padinhateeri,R. (2014) Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Res.*, **42**, 128–136.
28. Raghuraman,M.K., Winzeler,E.A., Collingwood,D., Hunt,S., Wodicka,L., Conway,A., Lockhart,D.J., Davis,R.W., Brewer,B.J. and Fangman,W.L. (2001) Replication dynamics of the yeast genome. *Science*, **294**, 115–121.
29. Gruss,C., Wu,J., Koller,T. and Sogo,J.M. (1993) Disruption of the nucleosomes at the replication fork. *EMBO J.*, **12**, 4533–4545.
30. Groth,A., Rocha,W., Verreault,A. and Almouzni,G. (2007) Chromatin challenges during DNA replication and repair. *Cell*, **128**, 721–733.
31. MacAlpine,D.M. and Almouzni,G. (2013) Chromatin and DNA replication. *Cold Spring Harbor Perspect. Biol.*, **5**, a010207.
32. Whitehouse,I. and Smith,D.J. (2013) Chromatin dynamics at the replication fork: there's more to life than histones. *Curr. Opin. Genet. Dev.*, **23**, 140–146.
33. Sogo,J.M., Stahl,H., Koller,T. and Knippers,R. (1986) Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures. *J. Mol. Biol.*, **189**, 189–204.
34. Radman-Livaja,M., Verzijlbergen,K.F., Weiner,A., van Welsem,T., Friedman,N., Rando,O.J. and van Leeuwen,F. (2011) Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biol.*, **9**, e1001075.
35. Rodriguez,J. and Tsukiyama,T. (2013) ATR-like kinase Mec1 facilitates both chromatin accessibility at DNA replication forks and replication fork progression during replication stress. *Genes Dev.*, **27**, 74–86.
36. Pospelov,V., Russev,G., Vassilev,L. and Tsanev,R. (1982) Nucleosome segregation in chromatin replicated in the presence of cyclohexamidine. *J. Mol. Biol.*, **156**, 79–91.
37. Annunziato,A.T. (2005) Split decision: what happens to nucleosomes during DNA replication? *J. Biol. Chem.*, **280**, 12065–12068.

38. Tran,V., Lim,C., Xie,J. and Chen,X. (2012) Asymmetric division of Drosophila male germline stem cell shows asymmetric histone distribution. *Science*, **338**, 679–682.
39. Smith,D.J. and Whitehouse,I. (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature*, **483**, 434–438.
40. Anderson,J.D. and Widom,J. (2000) Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.*, **296**, 979–987.
41. Mazurkiewicz,J., Kepert,J.F. and Rippe,K. (2006) On the mechanism of nucleosome assembly by histone chaperone NAP1. *J. Biol. Chem.*, **281**, 16462–16472.
42. Ransom,M., Dennehey,B.K. and Tyler,J.K. (2010) Chaperoning histones during DNA replication and repair. *Cell*, **140**, 183–195.
43. Annunziato,A.T. (2012) Assembling chromatin: the long and winding road. *Biochim. Biophys. Acta*, **1819**, 196–210.
44. Liu,W.H. and Churchill,M.E.A. (2012) Histone transfer among chaperones. *Biochem. Soc. Trans.*, **40**, 357–363.
45. Korber,P., Luckenbach,T., Blaschke,D. and Horz,W. (2004) Evidence for histone eviction in trans upon induction of the yeast PHO5 promoter. *Mol. Cell. Biol.*, **24**, 10965.
46. Boeger,H., Griesenbeck,J., Strattan,J.S. and Kornberg,R.D. (2004) Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Mol. Cell.*, **14**, 667–673.
47. Korber,P. (2012) Active nucleosome positioning beyond intrinsic biophysics is revealed by in vitro reconstitution. *Biochem. Soc. Trans.*, **40**, 377–382.
48. Schiessel,H. (2003) The physics of chromatin. *J. Phys.: Condens. Matter*, **15**, R699–R774.
49. Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
50. Yan,J., Maresca,T.J., Skoko,D., Adams,C.D., Xiao,B., Christensen,M.O., Heald,R. and Marko,J.F. (2007) Micromanipulation studies of chromatin fibers in *Xenopus* egg extracts reveal ATP-dependent chromatin assembly dynamics. *Mol. Biol. Cel*, **18**, 464–474.
51. Möbius,W. and Gerland,U. (2010) Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS Comput. Biol.*, **6**, e1000891.
52. Chereji,R.V. and Morozov,A.V. (2011) Statistical mechanics of nucleosomes constrained by higher-order chromatin structure. *J. Stat. Phys.*, **144**, 379–404.
53. Mejlvang,J., Feng,Y., Alabert,C., Neelsen,K.J., Jasencakova,Z., Zhao,X., Lees,M., Sandelin,A., Pasero,P., Lopes,M. *et al.* (2014) New histone supply regulates replication fork speed and PCNA unloading. *J. Cell Biol.*, **204**, 29–43.
54. Zhang,Z., Wippo,C.J., Wal,M., Ward,E., Korber,P. and Pugh,B.F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, **332**, 977–980.
55. Poli,J., Tsaponina,O., Crabbé,L., Keszthelyi,A., Pantesco,V., Chabes,A., Lengronne,A. and Pasero,P. (2012) dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.*, **31**, 883–894.

Supplementary Text

I. DIFFERENT CHOICES FOR THE DEPENDENCE OF THE KINETICS ON THE NUCLEOSOME-NUCLEOSOME INTERACTION

In general, the nucleosome assembly and eviction reactions within our model could both be configuration-dependent, leading to the position-dependent rates $\tilde{r}_+(x)$ and $\tilde{r}_-(x)$. However, the ratio between these reactions must be consistent with the Boltzmann factor at every position x,

$$\frac{\tilde{r}_+(x)}{\tilde{r}_-(x)} = \frac{r_+}{r_-}\,\mathrm{e}^{-V/k_BT} \tag{1}$$

where V is the net change in interaction potential produced by the addition of a nucleosome at position x (taking into account all contributing nearest-neighbour interactions[1]). Therefore, the most general form for the effective rates is

$$\begin{aligned} \tilde{r}_+(x) &= r_+\,\mathrm{e}^{-\theta V} \\ \tilde{r}_-(x) &= r_-\,\mathrm{e}^{(1-\theta)V}, \end{aligned} \tag{2}$$

where the parameter θ is in the range between zero and one. Values of θ close to one indicate rate-modulation at the point of assembly, while values close to zero indicate rate-modulation at the point of eviction. The nucleosome density as a function of time for various θ values is plotted in Fig. S1 for the three values of r_+ considered in the main text (as in the main text, we use time units such that $r_- = 1$).

[1] The main text Eq. 2 implicitly assumes that the nucleosomes immediately to the left and right of the assembly site x do not already overlap. In the extremely rare case of such a previously existing overlap, V is the sum of the interactions between the new nucleosome and its neighbors less the preexisting interaction between those neighbors.

II. GAP DISTRIBUTION AT THE ONSET OF CRAMMING

When SoNG adsorption reaches the cramming density of $\rho_c = 4.48 \cdot 10^{-3}$ nucleosomes/bp, the density, per bp, of gaps with size x is given by

$$\begin{aligned} G_c(x) &= A\, \mathrm{e}^{f(x)} \\ f(x) &= \begin{cases} -v(x) & \text{if } x < a \\ \alpha_1 x + b_1 & \text{if } a \le x < x^* \\ \alpha_2 x + b_2 & \text{if } x^* \le x \end{cases} . \end{aligned} \tag{3}$$

where $v(x)$ is the interaction potential from the main text, α_1, b_1 and α_2, b_2 are best fit values taken from regression over the intervals $[a, 2a)$ and $[2a, 3a)$ respectively, and x^* is the 'corner point' at which the exponential slopes intersect. A is a normalization constant, such that $\sum_x G_c(x) = \rho_c$, the number of nucleosomes per bp, (or, equivalently, the number of gaps per bp). Fig. S3 shows a histogram of gap sizes for our three time series at the onset of cramming, and Table S1 shows the values of the above parameters for each case.

III. SIMULATION PARAMETERS

In Table S2 we list the detailed parameters used in the Monte Carlo simulations for all figures of the main text and the supplement. Figure S5 is a reproduction of the corresponding Fig. 4 in the main text, with detailed simulation parameters added.

Supplementary Figures

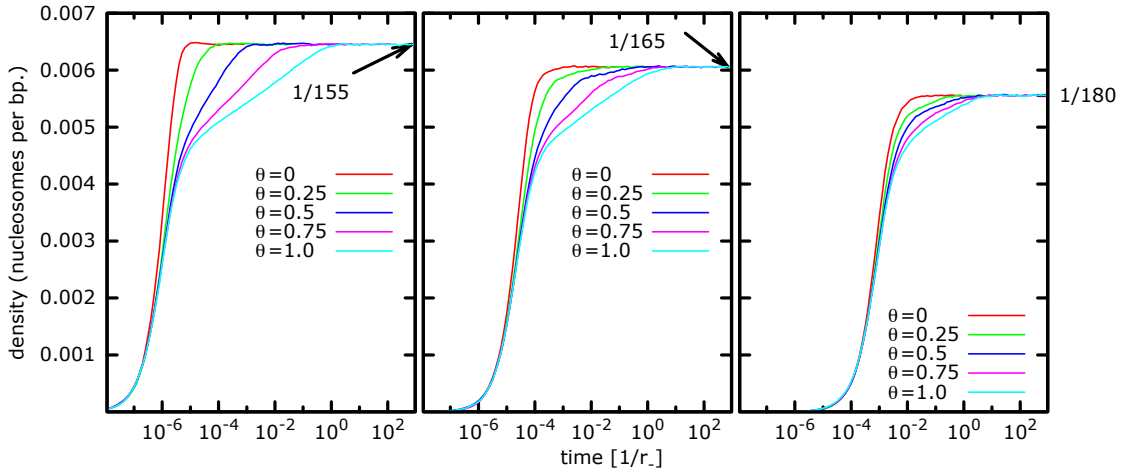


FIG. S1: Equilibration to our three final average densities under consideration for different values of θ, corresponding to different applications of the Boltzmann factor in the simulations. At $\theta = 1$ (the choice always presented in the main text), the speed-up of the filling process, relative to the HaNG model, is least pronounced. Thus, the choice of $\theta = 1$ is conservative with respect to our main conclusion.

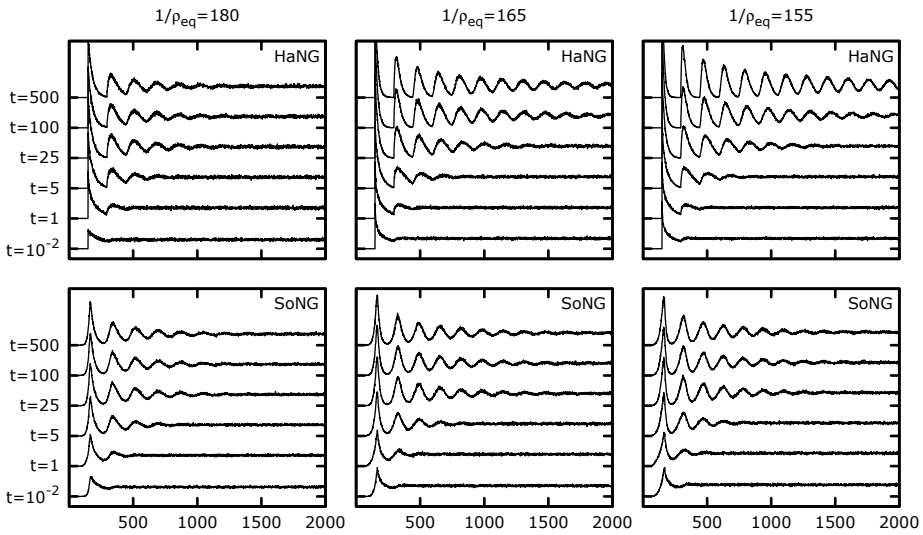


FIG. S2: Dynamics of nucleosome pattern formation close to a barrier (a fixed nucleosome) within the HaNG and SoNG model with three different steady-state densities (linear DNA with one nucleosome fixed at $x = 0$). The highest density case of the HaNG model has not reached the steady state even at $t = 500$. In the SoNG model, the patterns equilibrate at roughly the same time of about $t = 25$ for all three final densities. The equilibration times in the HaNG model, on the other hand, strongly depend on the final density (see also main text).

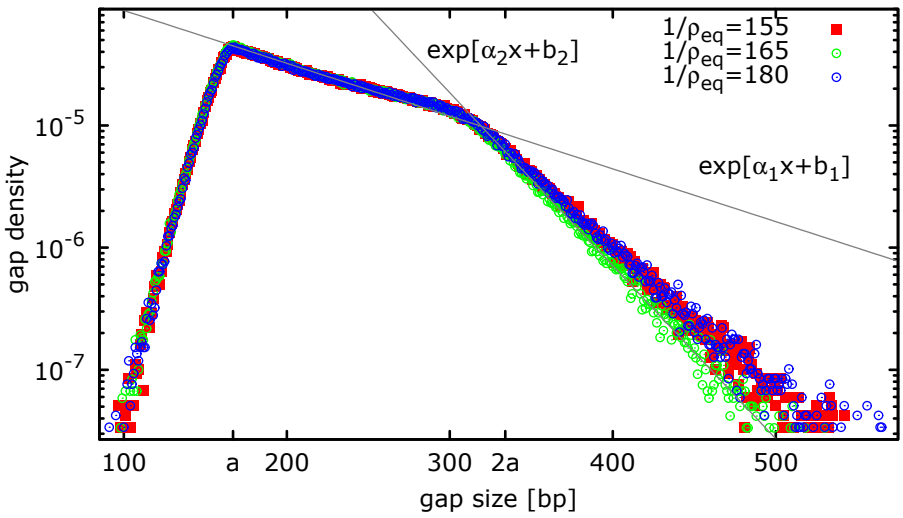


FIG. S3: Gap distribution at the onset of cramming for our three time-series. The nucleosome footprint $a = 167$ as well as $2a = 334$, is shown on the x-axis, roughly indicating points where the slope changes. For $x < a$, the distribution follows Boltzmann statistics for the energy of interaction between neighbours, while for $a \leq x < 2a$, and for $x > 2a$, the distributions are exponential. Regression lines from the latter two regions intersect at $x^* \lesssim 2a$, and the parameters describing this distribution are listed in Table S1.

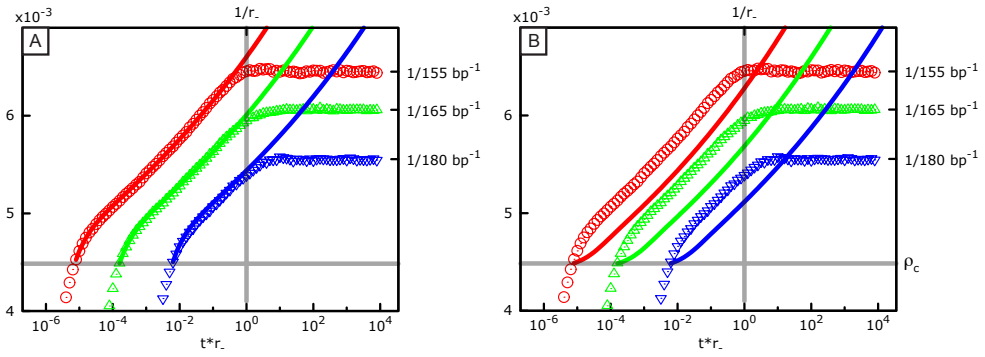
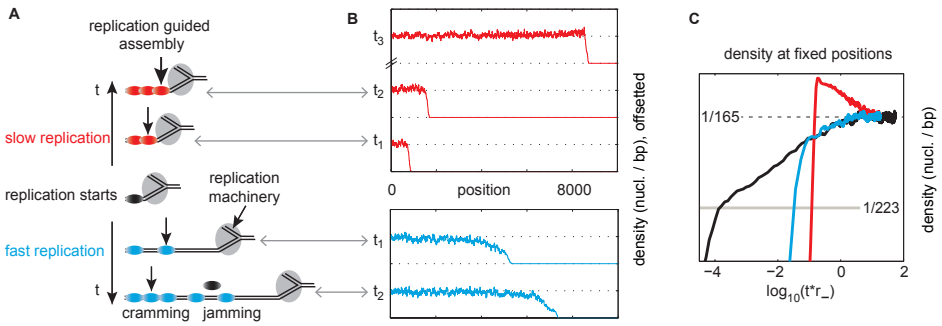


FIG. S4: Relative contributions of the three regimes of gap sizes to the cramming dynamics shown in Fig. 3B of the main text. (A) Only gaps of size larger than the nucleosome footprint a are allowed to fill, i.e. only regimes (ii) and (iii) of Fig. 3A are included in the calculation of $\rho(t)$ using Eq. 4 of the main text (solid lines). The result describes the simulated data (symbols) equally well as in Fig. 3B, indicating that the contribution of regime (i) is insignificant. (B) Only regime (ii) is included. In this case, the calculated curves no longer describe the initial rapid filling in the simulation data. This indicates that regime (iii) is responsible for this initial behavior. However, regime (ii) suffices to explain the dominant behavior, i.e. the slope of the filling curves.



common parameters:

number of initially naked bp: 20*167 = 3340

total number of bp: 80*167 = 13360

r+ = 221 /bp/s
r- = 1 /s

number of runs averaged: 30
(plus 2000 consecutive bp for panel C)

simulation time: 0 s to exp(4) s

slow replication parameters:

speed of right boundary: 1485 bp/s

times of snapshots in (b):
0.64 s
1.18 s
5.92 s

positions for time trace in (c):
poss 10,000 - 12,000
(averaged)

fast replication parameters:

speed of right boundary: 37128 bp/s

times of snapshots in (b):
0.15 s
0.20 s

positions for time trace in (c):
poss 10,000 - 12,000
(averaged)

FIG. S5: Detailed simulation parameters used for the illustration of the wake filling mechanism in Fig. 4 of the main text (the figure is reproduced here for ease of reference).

Supplementary Tables

$1/\bar{\rho}$[bp]	$t_c[1/r_-]$	α_1	α_2	b_1	b_2	$x*$	A
155	6.3 x10^{-6}	-9.6 x10^{-3}	-2.8 x10^{-2}	1.6	7.5	316	4.5 ×10^{-5}
165	1.6 x10^{-4}	-1.0 x10^{-2}	-3.3 x10^{-2}	1.7	8.9	319	4.6 ×10^{-5}
180	5.0 x10^{-3}	-9.5 x10^{-3}	-2.7 x10^{-2}	1.5	7.2	314	4.4 ×10^{-5}

TABLE S1: Constants describing the distributions in figure S3 for the three time series. Each curve in Fig. 2 reaches the onset of cramming at time t_c, at which point the probability distribution of inter-nucleosome gaps can be characterized by a piecewise exponential function with parameters shown here.

figure	panel / trace	geometry	system size L	number of runs
2	A,B	periodic	6000	400
	C,D	periodic	6000	60000
	C trace 1/155	periodic		
	F	periodic	6000	60000
3	A	periodic	6000	20000
4	see parameters in Fig. S5			
5	B,C	linear	200 × particle size	10
S1		periodic	6000	400
S2	each color (final density): see parameters for Fig. 3			
S3		periodic	8000	60000

TABLE S2: List of simulation parameters. For Fig. 5, the system size is given in multiples of the particle size, which is 147 for HaNG and 167 for SoNG.

5.2 A Statistical Physics Perspective

The study included above was motivated by the biological application of 1-D adsorption of histones on DNA, and the results have particular relevance in light of ongoing research in the kinetics of chromatin assembly[111]. It turns out, however, that this application serves as an impetus for the study of more fundamental properties of non-equilibrium physical processes which we explore now. The second manuscript attached at the end of this chapter presents the major findings of this more general study and, as of July 28 2015, has been accepted by peer-review for publication. The supplementary materials of this text (denoted reference 21 therein) consist primarily of content presented explicitly in this chapter, though to facilitate ease of reading, only portions of the text from this supplement are used here, rather than the whole document.

Reversible one-dimensional sequential adsorption of hard-exclusive particles, and the corresponding equilibration process in the above-jamming regime, has been an area of active research for some time[112, 113]. The problem, in addition to demonstrating intriguing physics in its own right, exhibits the same density dynamics as granular media [114, 115, 116, 105], and presents intriguing applications to related problems[117, 118, 119].

Part of the reason this problem is interesting is that it exhibits partially broken ergodicity[120] in the sense that during unjamming, only a restricted portion of the state-space is accessible. Various approach laws have been proposed to describe this relaxation process, such as using mean-field theory to describe the void distribution from the density[121, 120], or via phenomenological means using a stretched exponential[122]. Up to this point, however, available literature on the kinetics of line filling has been restricted to particles that interact via exclusion. The only study on kinetic filling with finite interactions we are aware of relied on stochastic interaction upon overlap (i.e. a potential that is either zero or infinite randomly)[123] -a very different form of 'soft' interaction from the deterministic potential we are interested in. To explore this topic in a more generic way, rather than the nucleosome-specific potential given in Eq. 2.5, we use the linear interaction potential

$$\varphi(x) = \varepsilon(1 - x) \qquad \text{for} \qquad x < 1. \tag{5.5}$$

which defines an effective particle size of unity. The filling dynamics observed for particles with an interaction potential from Eq. 5.5 are qualitatively the same as for the SoNG potential derived in Chapter 2, however this form is intended to be archetypal for other applications with interactions that can be roughly linearized. Using the same Monte Carlo simulations described in Section 2.5, the density of particles as a function of time for various levels of stiffness ε are shown in Fig. 5.3. While this interaction potential served purely to diminish adsorption rates (rather than expedite desorption,) alternative implementations of detailed balance were discussed in the previous document included above.

Note that for the upper curve in Fig. 5.3 the density trace is non-monotonic, reaching a transient maximum we refer to as the 'overshoot' shortly before $tr_- \approx 1$ and then relaxing *downward* toward equilibrium. This behavior is reminiscent of other transient density profile maxima which have been shown in 2D adsorption experiments[119].

The three curves shown illustrate the emergence of this behavior with decreasing ε, and a

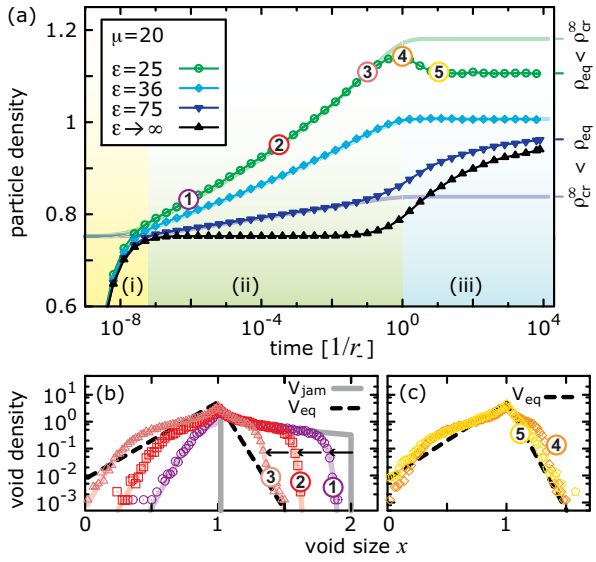


Figure 5.3: (a) Time traces of $\rho(t)$ for constant r_+, r_-, assuming a linear interaction potential as given in Eq. 5.5, with varying interaction strengths ε showing the transition into the overshoot regime. The void distribution at time points $1-5$ indicated in (a) are shown in (b),(c). Overlaid on the stochastic data (symbols) are transparent lines determined using a more sophisticated cramming formula than what was used in the previous manuscript. This theory provides the transparent density traces in (a) using Eq. 5.13, as well as the transparent void distributions in (b) using Eq. 5.12. The criteria that the asymptote of this cramming density exceed the equilibrium value delineates the regime where a transient maximum is observed.

more exhaustive exploration of the phase-space for the overshoot is presented in Fig. 3 of the manuscript attached below.

The cause of this behavior is not obvious, and to understand it, a first step is to see whether the same behavior is observed in a calculation of the density using mean-field void probabilities, since non-mean-field effects have been shown to play an important role in equilibration for the case of hard particles[124].

The same mean-field description that was applied to the jammed configuration of hard particles used in Eq. 5.1 has also been used to include time dynamics by treating voids of size x as dynamic variables $V(x,t)$[125]. In so doing, the evolution of the density in time can be

predicted. We extend this reasoning one step further to include the above neighbor interaction

$$\begin{aligned}\frac{\partial}{\partial t}V(x,t) = &- r_+ V(x,t)\int_0^x \mathrm{d}y\, e^{-\varphi(y)-\varphi(x-y)+\varphi(x)} \\ &+ 2\,r_+\int_x^\infty \mathrm{d}y\, V(y,t)\, e^{-\varphi(x)-\varphi(y-x)+\varphi(y)} \\ &- 2\,r_- V(x,t) \\ &+ \frac{r_-}{\rho(t)}\int_0^x \mathrm{d}y\, V(y,t)V(x-y,t)\ , \end{aligned} \tag{5.6}$$

where $\rho(t) = \int V(x,t)\,dx$ owing to the one-to-one correspondence between particles and voids.

In actual calculation, we coarse-grain Eq. 5.6, into the form of discrete $k-$mers along a lattice of finite size L with periodic boundary conditions, with k and L chosen large enough to approach an effective quasi-continuum, as shown below.

While the length variable x, in parenthesis, is used in Eq. 5.6 to denote separation between neighbors in units of the particle size, it is useful to employ subscript indices to denote the discrete number of intervening lattice sites between the positions of neighboring particles along this lattice; a circumflex is also used to denote quantities measured with respect to discrete lattice site lengths rather than particle size. Hence, V_0 refers to immediately adjacent lattice positions with maximal overlap interaction φ_0, while a V_{k-2} void carries the minimum non-zero interaction energy, $\varphi_{k-2} = \hat{\varepsilon}$, where $\hat{\varepsilon} = \varepsilon/k$. Likewise $\varphi_m = (k-m-1)\varepsilon/k$ for $0 \le m \le (k-1)$ (zero for $m \ge k$). Finally, the on-rate per lattice site, $\hat{r}_+$, is related to the on-rate per footprint r_+ by $r_+ = k\hat{r}_+$ as with the number of particles per footprint $\rho = k\hat{\rho}$. The non-interacting chemical potential cited refers to $\mu_0 = \ln(r_+) = \ln(k\hat{r}_+)$. The discrete analogue of Eq. (5.6) is then

$$\dot{V}_{\mathrm{m}}(t) = \begin{cases} -2r_- V_m(t) \\ +2\hat{r}_+ \sum_{i=m+1}^{L-1} V_i(t) e^{-\varphi_m - \varphi_{i-1-m} + \varphi_i} \\ -\hat{r}_+ V_m(t) \sum_{i=0}^{m-1} e^{-\varphi_i - \varphi_{m-1-i} + \varphi_m} \\ +\frac{r_-}{\rho^*} \sum_{j=0}^{m-1} V_j V_{m-1-j}(t), \\ \\ (\text{if } 0 \leq m < L-2) \\ \\ \hline \\ -r_- V_m(t) \\ +\hat{r}_+ L V_{m+1}(t) e^{-\varphi_m} \\ -\hat{r}_+ V_m(t) \sum_{i=0}^{m-1} e^{-\varphi_i - \varphi_{m-1-i} + \varphi_m} \\ +\frac{r_-}{\rho^*} \sum_{j=0}^{m-1} V_j V_{m-1-j}(t), \\ \\ (\text{if } m = L-2) \\ \\ \hline \\ -\hat{r}_+ L e^{-\varphi_{m-1}} V_m(t) + r_- V_{m-1}(t) \\ \\ (\text{if } m = L-1)\,. \end{cases} \tag{5.7}$$

Note the parameter $\rho^* \neq \rho$; for arbitrarily large but finite systems, the finite number of particles necessitates a correction to Eq. (5.6) to maintain conservation of space. Applying this conservation relation to Eq. 5.7 yields

$$\rho^*(t) = \frac{\sum_{n=0}^{L-1} (1+n) \sum_{j=0}^{n-1} V_{\mathrm{j}}(t) V_{n-1-j}(t)}{2 \sum_{m=0}^{L-2} (1+m) V_m(t)}, \tag{5.8}$$

while the ratio $\rho^*(t)/\rho(t)$ converges to unity quickly as an initially empty system fills. The density of particles, (our main observable of interest) is then taken as the number of particles per k lattice sites, ρ.

Note in Eq. 5.7 the two characteristic time-scales of adsorption and desorption $r_\pm$, which set the onset of jamming and equilibration respectively. As seen in Fig. 5.4 the $\rho(t)$ predicted by Eq. 5.7 is very much consistent with the values obtained from full stochastic simulation. Moreover, as shown in Fig. 5.5, discretizing the continuous particles into $k-$mers has little

effect on ρ, provided k is chosen sufficiently fine. Fig. 5.5 suggests that $k \gtrsim 10$ is sufficient to be regarded as an effective quasi-continuous particle. For the remainder of this text, unless otherwise stated, quasi-continuous data was gathered by using $k = 50$.

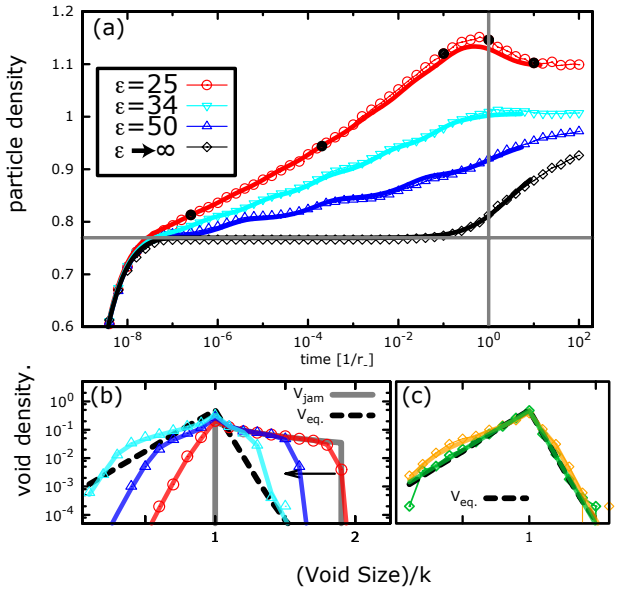


Figure 5.4: (a) density traces during the filling process using Monte Carlo data (symbols) as well as the mean-field void equations from Eq. 5.7 (solid lines with corresponding color-code) for k=10, $\mu_0 = 20$, and varying ε; a small but significant discrepancy can be observed near $t \approx 1/r_-$. (b),(c) void density distributions before and after $1/r_-$ respectively, using Monte Carlo data (symbols) and profiles taken from Eq. 5.7 (semi-transparent lines). profiles in (b),(c) are taken at times indicated by black dots in (a).

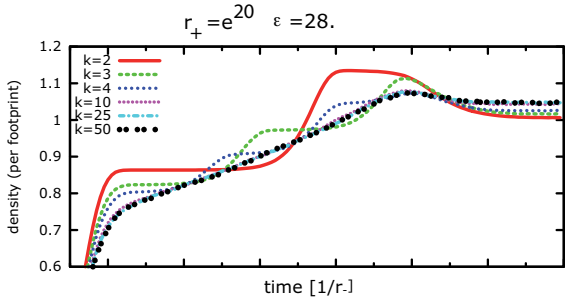


Figure 5.5: Filling dynamics for various levels of coarse-graining assuming constant on-rates per particle size, and interaction strengths ε. For measuring ρ, $k \gtrsim 10$, seems to be sufficiently fine to reproduce continuum properties.

From the above, it can be said that the qualitative features of filling, cramming and relaxation seen in 5.3 seem to be relatively insensitive to coarse-graining or mean-field approximations. A natural next step in understanding the overshoot comes by revisiting the cramming analysis given in Eq. 5.4.

5.3 Cramming with Finite Density Saturation

Eq. 5.4 makes the rather simplistic assumption that all binding is irreversible. While this assumption is useful in analyzing system behavior well before the formation of stacked arrays, it also unfortunately overestimates the density by breaking detailed balance locally. Assuming each of the voids within Eq. 5.1 can be crammed -now reversibly- against some repulsive interaction potential, the probability of voids of size x remaining unfilled at time t, $P(x,t)$, follows the relation:

$$\frac{d}{dt}P(x,t) = -P(x,t)\gamma(x) + (1 - P(x,t))r_{-}. \tag{5.9}$$

Here, the second term is the probability that the void has been filled, multiplied by the rate of desorption. From another perspective, the two terms in Eq. 5.9 reflect the first and last terms of Eq. 5.6. During the cramming period, $t \in [1/r_{+}, 1/r_{-}]$, the opportunity to fill very large voids has already been exhausted and the bounding particles on either side are not yet departing, thus the second and third terms of 5.6 can be neglected.

This partial relaxation of irreversibility for the cramming particles has the somewhat arbitrary implication of treating particles that arrived before jamming as fixed in place, while those adsorbing during cramming do so reversibly (that is to say, the fourth term in Eq. 5.6 is effectively retained. The goal here is to model saturated filling on a local level, before many-body rearrangements. At any rate, such an estimate is a better approximation than the irreversible 'ratchet-like' behavior of Eq. 5.4, and we will see below that it actually produces reasonable asymptotic behavior that helps explain the overshoot. The solution to equation 5.9 supplies the elements of our diagonal probability matrix P_{xx} which, again, describes the likelihood of a gap of size x remaining unfilled:

$$\hat{P}(t)_{xx'} = \frac{\delta_{xx'}}{\gamma(x) + r_{-}} \left(r_{-} + \gamma(x) e^{-(\gamma(x)+r_{-})t} \right). \tag{5.10}$$

Likewise, the transition matrix $\hat{T}_{xx'}$ characterizes the creation of voids of size x from the filling of a void of size x':

$$\hat{T}_{xx'} = \begin{cases} 2\frac{e^{-\varphi(x)-\varphi(x'-x)+\varphi(x')}}{\gamma(x')} & \text{if } x < x' \\ 0 & \text{else .} \end{cases} \tag{5.11}$$

Note that $\int_x dx\, \hat{T}_{xx'} = 2$, since each each cramming reaction creates a net increase of one void. Taken together with a vector interpretation of void distributions V, V_{jam}, equations 5.10 and 5.11 suffice to give us the cramming distribution of voids of size x at time t

$$V(x,t) = \left(\hat{P}(t)_{xx} \cdot V_{\text{jam}}(x) + \hat{T}_{xy} \cdot (I - \hat{P}(t)_{yy}) \cdot V_{\text{jam}}(y) \right). \tag{5.12}$$

Snapshots of this void distribution are overlaid in transparency in Fig. 5.3(b) as described in the caption. The density of such a 'glassy' phase is

$$\rho_{\rm cr}(t) = \rho_{\rm jam} + \int dx V_{\rm jam}(x) \frac{\gamma_+(x)\left(1 - e^{-[\gamma_+(x)+r_-]t}\right)}{\gamma_+(x) + r_-}, \tag{5.13}$$

which is also overlaid, in transparency, on stochastic data in Fig. 5.3(a). Note that Eq. 5.13 reaches asymptotic behavior not seen in the projections from Eq. 5.4. Furthermore, the asymptotic level depends on the interaction potential; as a result, the overshoot in Fig. 5.3 becomes visible when $\rho_{\rm cr}$ yields a steady state result that exceeds the ultimate equilibrium density. This brings us to the basic premise of our explanation for the transient overshoot seen in Fig. 5.3: the void distribution established at the conclusion of RSA can be crammed to densities that are unsustainable when full freedom of redistribution is allowed for the many particles in the system.

To test this, we apply the same cramming theory in Eq. 5.13 under a *different initial condition.* Namely, one that will proceed through the cramming process directly into the equilibrium void distribution –a 'guiding' distribution, so to speak. For an equilibrium state with density $\rho_{\rm eq}$, and void density $V_{\rm eq}(x)$, the corresponding probability distribution of voids $P_{\rm eq}(x) = V_{\rm eq}(x)/\rho_{\rm eq}$ is obtained by simply normalizing $V_{\rm eq}(x)$ to unity. If one were to remove alternating particles from this arrangement, the distribution of 'double-voids' would then be

$$P_2(x) = \int_0^x dx' \frac{V_{\rm eq}(x')}{\rho_{\rm eq}} \frac{V_{\rm eq}(x-x')}{\rho_{\rm eq}}. \tag{5.14}$$

Using this, we consider a 'guiding' initial configuration of particles obtained by removing isolated particles from the equilibrium configuration until the density is reduced to $\rho_{\rm jam}$. The void density distribution for this guiding configuration V_g is mixed from the distributions $P_2(x)$ and $P_{\rm eq}(x)$ in such a ratio as to prepare the system in the same initial density, $\rho_{\rm jam}$, as before:

$$V_g(x) = (\rho_{eq} - 2(\rho_{eq} - \rho_{\rm jam}))\, P_{eq}(x) + (\rho_{eq} - \rho_{\rm jam}) P_2(x). \tag{5.15}$$

Thus, the 'guided' density evolution follows

$$\rho_{\rm g}(t) = \rho_{\rm jam} + \int dx V_{\rm g}(x) \frac{\gamma_+(x)\left(1 - e^{-[\gamma_+(x)+r_-]t}\right)}{\gamma_+(x) + r_-}, \tag{5.16}$$

which is identical to Eq. 5.13, except that the initial condition that serves as a template for cramming is now V_g instead of $V_{\rm jam}$.

The importance of this seemingly subtle difference is made clear in Fig. 5.6. Here, cramming of V_g proceeds directly to equilibrium without any high-density intermediate. On its own, this observation is rather trivial (each removed particle is simply crammed back into its original position); however when juxtaposed against the density evolution using $V_{\rm jam}(x)$ as a template, it shows that the initial condition of the cramming process is what determines the asymptotic density until many-particle rearrangements at timescales $t > 1/r_-$.

To reiterate, the overshoot is a consequence of the dependence of the asymptotic density on the initial condition subject to cramming, and how this density compares to the ultimate equilibrium density once many-particle rearrangements take effect. Under the right conditions, this asymptotic density is higher than what can be sustained at equilibrium when many-particle rearrangements are taking place, leading to a density 'collapse' near time scales $t \approx 1/r_-$, and an apparent overshoot.

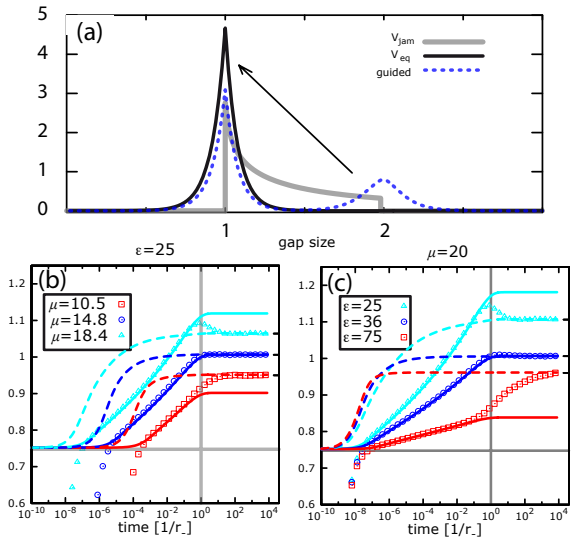


Figure 5.6: (a) Initial distributions V_{jam} and V_{guided} compared to the equilibrium distribution V_{eq}. Cramming calculations from Eq. 5.13 and Eq. 5.16 respectively are then applied to these initial distributions under varying (b)μ_0 and (c)ε. In both cases, cramming of the guided distribution (dashed) proceeds directly to equilibrium (as expected), while the 'slow and steady' cramming of V_{jam} leads to different asymptotics that can, under the right conditions, exceed ρ_{eq}. Whenever this occurs, the Monte Carlo data show a transient maximum in the density profile.

5.4 Dimers

Earlier in this chapter we considered an array undergoing reversible adsorption of particles occupying k adjacent lattice sites, where k was set to either 147, or 165 for the HaNG and SoNG models respectively. These k values are well into the regime of quasi-continuum behavior where the discrete 'graininess' of the lattice sites in the array (i.e. DNA base pairs in the context of nucleosomes) is insignificant. The last section showed that k can be as low as 20, or even 10, while preserving much of the characteristics of a continuum.

Unfortunately, a full analytic description of the relaxation of these larger particles is not available. Nevertheless, a fairly clear qualitative understanding can be gleaned from a reduced system that *is* fully solvable. In the extreme coarse-graining limit, $k = 2$, dimer particles bind reversibly to an infinite array of lattice sites. As before, we focus on the more kinetically interesting regime $r_+ \gg r_-$, where the initially empty substrate first reaches a jammed configuration via RSA before relaxing towards equilibrium.

Assuming the dimers are 'hard', and prohibit overlap, the density (per footprint of two lattice sites) of the jammed configuration has been shown to be $\rho_{\text{jam}} = 1 - e^{-2}$[126]. At this point, neighboring dimers can be separated by either 2 or 3 lattice lengths but no more. We refer to the density of such intervening voids as V_1, and V_2 respectively, while 'soft' dimers admit the additional possibility of single overlaps denoted V_0 with energetic penalty $\hat{\varepsilon}$; a schematic is given in Fig. 5.7.

On the typical time scales at which such jammed configurations become established, large

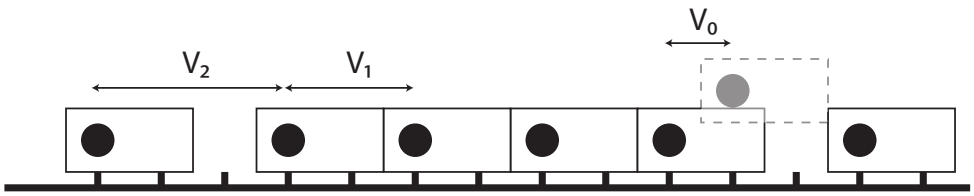


Figure 5.7: A schematic of a jammed configuration after RSA, with $V_{0,1,2}$ voids labeled. The transparent dimer is possible only upon cramming. As dimers are made increasingly stiff (i.e. as $\varepsilon \to \infty$), the concentration of overlaps V_0 diminishes, approaching the hard dimer limit, where $V_0 \to 0$. The convention of 'land-marking' dimers at their left position is arbitrary and done for clarity.

voids of size 3 or higher are extremely short lived. With this observation in mind, a truncation of the mean-field description given in Eq. 5.7, using only $V_0(t), V_1(t)$, and $V_2(t)$ as dynamic variables, can be obtained by tracing over larger gaps that act only as transient intermediates. The order of this approximation can be taken as the number of adjacent desorption events that are likely to happen in succession. The $n-$th order truncated mean-field dynamic equations are then:

$$\begin{aligned}
\dot{V}_0 = &\quad -2r_-V_0 - 2r_-\frac{V_0^2}{\rho} + 2\hat{r}_+\mathrm{e}^{[-2\hat{\varepsilon}]}V_1 + 2\hat{r}_+\mathrm{e}^{[-\hat{\varepsilon}]}V_2 + f_0^{(n)}(V_0,V_1,V_2)\\
\dot{V}_1 = &\quad 2r_-\frac{V_0^2}{\rho} - 2r_-\frac{V_0V_1}{\rho} - \hat{r}_+\mathrm{e}^{[-2\hat{\varepsilon}]}V_1 + 2\hat{r}_+\mathrm{e}^{[-\hat{\varepsilon}]}V_2 + f_1^{(n)}(V_0,V_1,V_2)\\
\dot{V}_2 = &\quad 2r_-\frac{V_0V_1}{\rho} - 2\hat{r}_+\mathrm{e}^{[-\hat{\varepsilon}]}V_2 + f_2^{(n)}(V_0,V_1,V_2)
\end{aligned} \tag{5.17}$$

where $\rho(t) = V_0(t) + V_1(t) + V_2(t)$ is the total particle density. The terms $f_{0-2}^{(n)}$ are generally the contributions to the dynamic variables from immediate refilling of transient states larger than V_2 after a maximum of n sequential desorption events; they are constrained to satisfy overall conservation of space. A 'first order' approximation ($n = 1$), for example would entail assuming that every desorption event is immediately followed by an adsorption that occurs –with equal probability– at any of the non-interacting sites that become available. An illustration is provided in Fig. 5.8. The additional dynamics to first order are then given by:

$$\begin{aligned}
f_0^{(1)} = &\quad 0\\
f_1^{(1)} = &\quad \frac{4r_-V_0V_2}{\rho} + \frac{4r_-V_2^2}{\rho}\\
f_2^{(1)} = &\quad -\frac{2r_-V_0V_2}{\rho} - \frac{8r_-V_2^2}{3\rho}
\end{aligned} \tag{5.18}$$

which account for transient states up to V_5 –the maximum transient void size at this order of approximation. A clear problem in this description arises when one considers that all possible reactions must either conserve or increase the total number of particles. The asymptotic density of hard particles will then always be precisely 1.0, and density increase is then monotonic *by construction.*

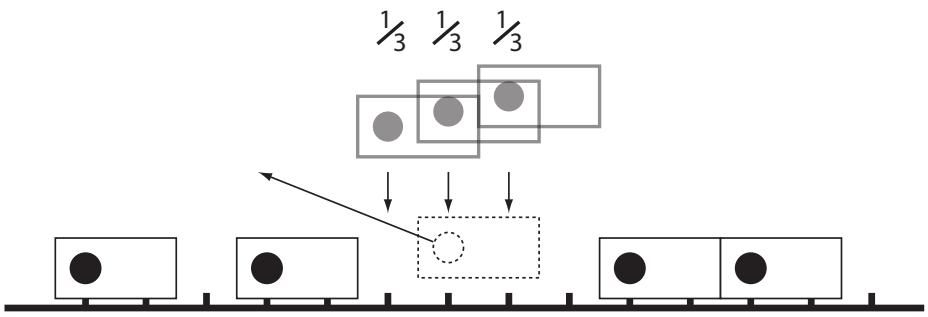


Figure 5.8: A desorbing dimer with V_2 voids on either side leaves 3 available sites for unobstructed adsorption. In our 'first order' treatment, each of the exposed sites is filled with probability $\frac{1}{3}$. Thus, with probability $\frac{2}{3}$, the incoming dimer lands to one side, and is immediately followed by an additional dimer on the other side, resulting in a net increase of 1 particle.

Thus, we proceed to a second order approximation which allows for at most 2 successive adjacent desorption events. At this order, the probability that the first desorption event is followed immediately by at least one filling adsorption (previously unity) is now P_1, while the probability that the first desorption event is followed quickly by the desorption of one of the two adjacent dimers is P_2. These two quantities are given by:

$$\begin{aligned} P_1 &= \frac{d\,\hat{r}_+}{d\,\hat{r}_+ + 2r_-} \\ P_2 &= \frac{2r_-}{d\,\hat{r}_+ + 2r_-} \end{aligned} \tag{5.19}$$

where d is the number of non-interacting binding sites available (e.g. $d = 3$ in the reaction depicted in Fig. 5.8). The largest intermediate voids is then V_8, but subsequent filling then proceeds along the same principle for all intermediate voids $V_3 - V_8$ (i.e. all non-interacting binding sites are equally likely, and are filled until exhausted). The more complicated relations at second order represent a much improved approximation:

$$
\begin{aligned}
f_0^{(2)} = & -2r_-\frac{V_0V_2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)V_0 - 2r_-\frac{V_1^2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)V_0 \\
& -2r_-\frac{V_1V_2}{\rho^2}\left(\frac{2r_-}{2\hat{r}_+ + 2r_-}\right)V_0 - 2r_-\frac{V_2^2}{\rho^2}\left(\frac{2r_-}{3\hat{r}_+ + 2r_-}\right)V_0 \\
f_1^{(2)} = & \ 4r_-\frac{V_0V_2}{\rho}\left(\frac{\hat{r}_+}{\hat{r}_+ + 2r_-}\right) + 4r_-\frac{V_2^2}{\rho}\left(\frac{3\hat{r}_+}{3\hat{r}_+ + 2r_-}\right) \\
& +2r_-\frac{V_0V_2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)[V_0 + V_1 + 2V_2] - 2r_-\frac{V_1^2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)[V_0 + V_1] \\
& +2r_-\frac{V_1V_2}{\rho^2}\left(\frac{2r_-}{2\hat{r}_+ + 2r_-}\right)\left[V_0 + \frac{21}{15}V_2\right] + 2r_-\frac{V_2^2}{\rho^2}\left(\frac{2r_-}{3\hat{r}_+ + 2r_-}\right)\left[2V_0 + \frac{21}{15}V_1 + \frac{24}{9}V_2\right] \\
f_2^{(2)} = & -2r_-\frac{V_0V_2}{\rho}\left(\frac{\hat{r}_+}{\hat{r}_+ + 2r_-}\right) - \frac{8r_-V_2^2}{\rho}\left(\frac{\hat{r}_+}{3\hat{r}_+ + 2r_-}\right) \\
& -2r_-\frac{V_0V_2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)\left[\frac{1}{3}V_1 + V_2\right] + 2r_-\frac{V_1^2}{\rho^2}\left(\frac{2r_-}{\hat{r}_+ + 2r_-}\right)\left[V_0 + \frac{2}{3}V_1\right] \\
& -\frac{4r_-V_1V_2}{\rho^2}\left(\frac{r_-}{2\hat{r}_+ + 2r_-}\right)\left[\frac{1}{3}V_0 + \frac{14}{15}V_2\right] - \frac{4r_-V_2^2}{\rho^2}\left(\frac{r_-}{3\hat{r}_+ + 2r_-}\right)\left[V_0 + \frac{14}{15}V_1 + \frac{16}{9}V_2\right].
\end{aligned}
\tag{5.20}
$$

When Eq. 5.20 is implemented into Eq. 5.17, numerical time evolution can be taken to obtain the density ρ. The result compares favorably with Monte Carlo data during the initial filling phase and at equilibrium, as shown in Fig. 5.9. The density of energy $U(t)$ stored in the interaction of the soft dimers with finite $\hat{\varepsilon}$ indicates the density of overlaps in time. However, there remains a clear disparity during the relaxation period for $t > 1$, a point which can be attributed to non-mean-field effects described later in this chapter. For the moment, we consider a heuristic argument as to why monotonicity emerges for interacting dimers only as $\varepsilon \to \infty$.

5.4.1 The Hard Dimer Limit

In the equations of motion from Eq. 5.17, 5.20, although there are three dynamic variables, any of the three can be inferred from the other two due to conservation of space. Hence, there are only two degrees of freedom. When the limit $\varepsilon \to \infty$ is taken, these mean-field equations of motion reduce to hard-core dimer kinetics and $V_0 \to 0$, eliminating a further degree of freedom. The system then becomes one-dimensional, and we can express either of the two dynamic variables in terms of ρ as

$$
\begin{aligned}
V_1 &= 3\rho - 1 \\
V_2 &= 1 - 2\rho.
\end{aligned}
\tag{5.21}
$$

By incorporating Eq. 5.20 into Eq. 5.17, setting $V_0 \to 0$, summing over $\dot{V}_1 + \dot{V}_2$, and simplifying, we obtain

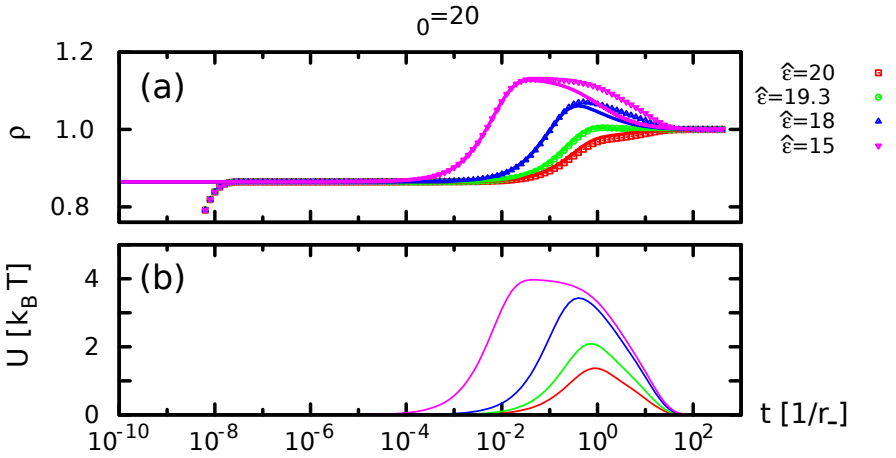


Figure 5.9: (a) Density of soft dimers as a function of time using stochastic simulation data (symbols) compared with the numerical results of Eq. 5.20 for various interaction strengths $\hat{\varepsilon}$; the second value $\hat{\varepsilon} = 19.3$ represents the marginal case where asymptotic cramming coincides with the equilibrium density. (b) The potential energy stored in the sum of neighbor interactions throughout the system U indicates the window of time over which overlaps occur. Initial filling density and equilibrium values from (a) are clearly in agreement, however during relaxation, deviation can be observed between the mean-field equations and full stochastic data.

$$\begin{aligned}\dot{\rho}(\rho) &= \frac{16r_-(1-\rho)^2}{\rho(3\tilde{r}_+ + 2r_-)}\left[\tilde{r}_+ + \frac{r_-}{\rho}\left(\frac{7(3\rho-2)}{15} + \frac{16(1-\rho)}{9}\right)\right] \\ &- \frac{4}{3}\frac{r_-^2(3\rho-2)^3}{\rho^2(\tilde{r}_+ + 2r_-)} + \frac{112}{15}\frac{r_-^2(3\rho-2)(1-\rho)^2}{\rho^2(2\tilde{r}_+ + 2r_-)}.\end{aligned} \tag{5.22}$$

and the utility of the description from Eq. 5.20 starts to become apparent. Expressing $\dot{\rho}$ as a function of ρ itself implies that this one-dimensional system is necessarily monotonic, as all one-dimensional systems must be. Recalling the definition $\hat{r}_+ = e^{[\mu_0]}$, we then plot this derivative for various μ_0 in Fig. 5.10.

Note the fixed points (where $\dot{\rho}(\rho) = 0$) in Fig. 5.10 converge at 1.0 in the limit $\mu_0 \to \infty$. The larger fixed point in the region $\rho > 1$ represents a non-physical solution, as hard dimers cannot exceed unity density; algebraic solution of the lower (stable) fixed point, however, yields the asymptotic density predicted by Eq. 5.17, which can then be compared to equilibrium calculations determined by the transfer matrix method. The resulting comparison agrees quite well for $\mu_0 \geq 3 - 4$ as shown in Fig. 5.11 (an additional trace in red, obtained from Eq. 5.27 is discussed below.)

Thus, the second-order truncated approximation to Eq. 5.7 for dimers reproduces not only the cramming and overshoot seen in stochastic simulations for soft interactions, but also the steady-state density for exclusive interactions. More importantly, it illustrates how the soft-dimers admit more complex dynamics than hard-dimers, and provides an indication as to why –at least at mean-field– monotonicity is violated only in the former case. Neither the state of the system nor the derivative of the density is uniquely defined by the density itself, and

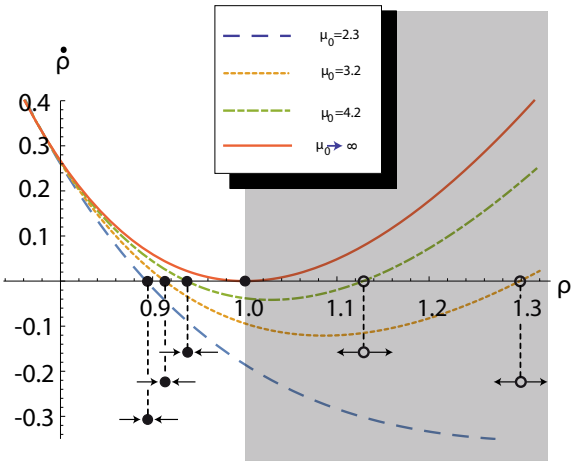


Figure 5.10: $\dot{\rho}$ plotted in terms of the density ρ itself for various μ_0 -such a reduction is only possible in the limit $\varepsilon \to \infty$. The zero intercepts for $\rho < 1$ represent fixed points of asymptotic density, while the shaded region represents non-physical solutions, since hard dimers are restricted to densities $\rho < 1$.

these other aspects of the system become particularly significant when particles are allowed to overlap.

While the mean-field description given here captures much of the dynamics, significant disparity with stochastic simulation during relaxation can nevertheless be observed in Fig. 5.9. Surprisingly, these deviations are not related to the truncation step made in Eq. 5.17, but rather are due to the assumption of mean-field itself. We elaborate on this in the following section.

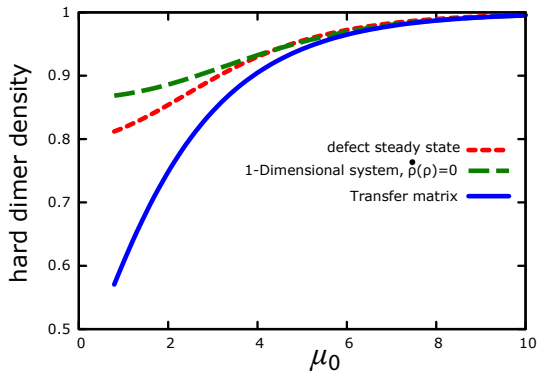


Figure 5.11: Steady state densities ρ of hard dimers determined exactly using the transfer matrix from chapter 2 (solid, blue), using the fixed point ρ defining $\dot{\rho} = 0$ in Eq. 5.22(dashed, green), and using the steady-state solution for defect creation and annihilation between dimers discussed below (red, dotted).

5.4.2 Non-Mean-Field Effects

The description above, starting from Eqs. (5.6) assumes that no correlations exist between the sizes of neighboring voids. We now consider an alternative picture, starting by referring to V_2 gaps between dimers as stacking 'defects', as they result in sub-optimal density arrangements.

When either particle adjacent to the defect desorbs and is replaced, the defect is shifted by two lattice sites half of the time. When two defects are adjacent, a new dimer will arrive to fill the space. Thus the behavior of these defects is equivalent to particle diffusion-annihilation, with the effective lattice for defects coarse-grained by a factor of 2. In fact, a solution to the diffusion-annihilation problem in 1-D is already available[125], and the parallel with filling dimers has already been drawn[127]. For an initially full lattice of particles which pair-wise annihilate irreversibly upon meeting one-another, the density is given by:

$$c(t) = I_0(2t)\mathrm{e}^{[-2t]} \tag{5.23}$$

where $I_0(x)$ is the modified Bessel function of the first kind with argument x, which at late times scales as $1/\sqrt{t}$.

The analogy here is not perfect since defects in the jammed state are never uniformly spaced. However, to capture the long-time dynamics we take the average density of defects in the jammed state $\frac{1}{\mathrm{e}^{[-2]}}$, as an approximate renormalized lattice. The unit of this array has spacing $\mathrm{e}^{[2]} \approx 7.389$, and the time required, on average, for one diffusing defect to reach its neighbor is then $\left(\frac{\mathrm{e}^{[2]}}{2}\right)^2 \approx 13.65$. The filling of hard dimers from the jammed state to complete coverage in the high-μ_0 limit is then approximated by

$$\begin{aligned} \rho(t) &= 1 - V_2(t) \\ &= 1 - I_0(2t')\mathrm{e}^{[-2t']} \\ &= 1 - I_0\left(\frac{e^4}{2} t e^{-\frac{e^4}{2}t}\right). \end{aligned} \tag{5.24}$$

Eq. 5.24 is not a significant advancement from what is already known[127]. What is, however, much more intriguing is that this same reasoning can apply even if the voids are crammed. Specifically, in the regime $\hat{r}_+ e^{-\varepsilon} > 1 > \hat{r}_+ e^{-2\varepsilon}$, V_2 gaps will be crammed by dimers that overlap with a single neighbor, however V_1 gaps will not, due their two-fold higher interaction penalty. By 'saturated' soft dimers, we refer to those in which V_2 gaps are generally filled, but not V_1.

Since each defect affords a crammed particle, in this scenario two defects sum to a net *gain* of one particle, as opposed to a loss in the hard case. Nevertheless, defect diffusion-annihilation still obeys they same kinetic principle and the same rate laws; the more general rule for ρ is then

$$\rho(t) = \begin{cases} 1 - I_0\left(\frac{e^4}{2}t\right)\mathrm{e}^{\left[-\frac{e^4}{2}t\right]} & \text{hard dimers} \\ 1 + I_0\left(\frac{e^4}{2}t\right)\mathrm{e}^{\left[-\frac{e^4}{2}t\right]} & \text{saturated soft dimers .} \end{cases} \tag{5.25}$$

Both of which approach unity as $t^{-1/2}$, up to the point where spontaneous pair-wise defect creation becomes relevant, as shown in Fig. 5.12.

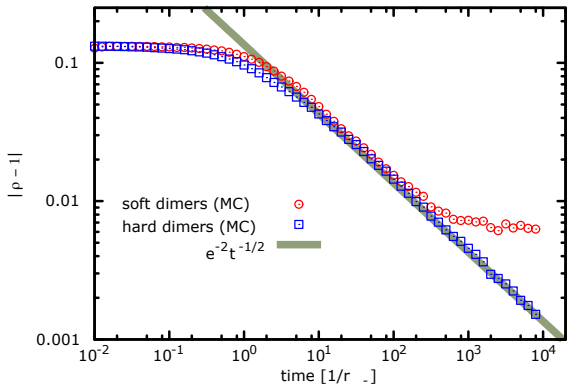


Figure 5.12: log-log plot of $|1-\rho|$ for full Monte-Carlo data shows the $t^{-1/2}$ dependence of both hard ($\mu_0 = 20, \hat{\varepsilon} \to \infty$) and soft ($\mu_0 = 20, \hat{\varepsilon} = 14$) dimers until steady state is approached via the finite rate of pair-wise defect creation.

The fact that the same relaxation behavior can be seen for both hard and saturated-soft particles for $k = 2$ brings us to an important question posed much earlier in this chapter: why do soft particles equilibrate so much faster in the continuous case? In other words: what remaining difference between dimers and continuous particles could account for the dramatic speed-up in equilibration for soft particles in the latter case?

The answer is guided readsorption of continuous soft particles. For continuous particles, the filling of voids between soft neighbors is not uniformly distributed throughout the gap, and the distribution of filling rates $\gamma(x)$ is strongly weighted near the center position $x/2$ for most cases that involve interaction. Thus, there is a natural preference for particles that desorb to be repositioned in an evenly-spaced manner. Hard-exclusive particles, on the other hand, readsorb randomly and uniformly throughout the available space. Close stacked positions then occur only rarely by chance, and so extended arrays form only very slowly.

In the case of dimers, however, there is no guiding energetic gradient for the readsorption –either for hard or soft dimers, since there is only a single interaction strength. As larger k particles are used, biased readsorption positioning becomes the determining factor in the faster equilibration for extended arrays of soft particles.

5.4.3 Steady State

The previous section demonstrated power-law approach in the density traces of dimers towards equilibrium, but only assuming that defect annihilation is irreversible.

In fact, steady-state pair-wise defect creation can also be observed for any finite $\hat{r}_+/r_-$. This, again, has been found in analogous problems such as spin-Ising lattices where 'defects' are boundaries between domains of parallel spin. These systems have been studied[125],

and the balance between annihilation and creation of defects is determined by the rate h with which boundary-pairs are created when an isolated spin within a domain flips. In our case, this rate is the product of r_-(since there is one dimer for every footprint in this limit) multiplied by $2r_-/(\hat{r}_+ + 2r_-)$ (the chance of an adjacent desorption event occurring before re-adsorption), and again by 1/3 (the probability of the replacement dimer landing in the middle). A schematic illustration of this is provided in Fig. 5.13.

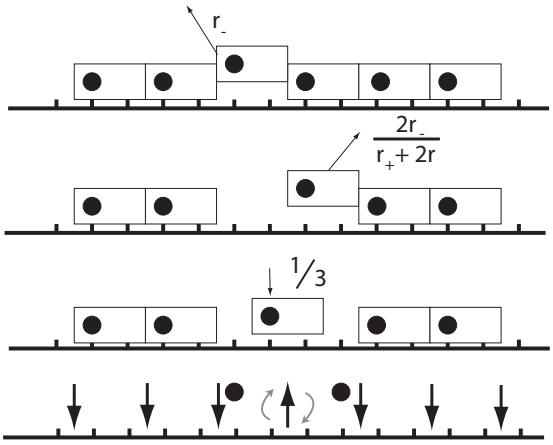


Figure 5.13: Creation of stacking defects, or, equivalently, domain boundaries in the 1-D spin Ising model

The defect-pair creation rate is then given as

$$h = \frac{r_-^2}{3(\hat{r}_+ + 2r_-)}. \tag{5.26}$$

For a given creation rate, the pair-wise creation/annihilation steady-state density of particles has been solved[125]:

$$\rho_{eq} = \frac{1}{2}\left(1 - \frac{\sqrt{h}}{1+\sqrt{h}}\right). \tag{5.27}$$

with h depending on $\mu_0 = \ln\left(\frac{\hat{r}_+}{r_-}\right)$ as given in Eq. 5.26. This equilibrium density, as a function of μ_0 is what is plotted in red in Fig. 5.11 against the exact result determined using the transfer matrix, in addition to the steady-state fixed point of $\dot{\rho}(\rho) = 0$ determined from Eq. 5.22, providing one further confirmation of our previous steady-state analysis.

The above discussion of dimer filling is intended to be a minimal example of the behavior of continuous particles that is fully solvable yet preserves the qualitative features of the latter: jamming; relaxation through collective rearrangements; cramming of soft particles predicted by 5.13, and its corresponding asymptotics; and overshoot, or non-monotonicity only for soft interacting particles. The key difference is that in the case of dimers, relaxation is no faster for soft particles than for hard, because there is no biasing energetic gradient guiding the readsorption of soft dimers (as there is for larger k particles), since interaction only occurs via single overlap.

5.4.4 Dimer Nonequilibrium Thermodynamics

As a final observation, by substituting dimers for extended particles and using explicit Monte Carlo statistical sampling, it becomes possible to study one-dimensional filling kinetics at another level of description. Namely, by employ an information theory approach one can study the nonequilibrium thermodynamics of filling. For an array of finite L sites, where indistinguishable soft dimers adsorb randomly, there is a total of 2^L unique arrangements of particles, and to each configuration i we attribute a probability p_i. For extended particles with large L, the number of configurations is too large for practical statistical sampling, but for dimers on a limited lattice we can use the Shannon measure of entropy[128]:

$$S(t) = -\frac{1}{L}\sum_n p_n(t)\ln\left[p_n(t)\right]. \tag{5.28}$$

We may also use the Kullback-Leibler divergence

$$H(t) = \frac{1}{L}\sum_n p_n(t)\ln\left[(p_n(t)/p_n^{\mathrm{eq}})\right] \tag{5.29}$$

to define the relative entropy $H(t)$ with respect to the equilibrium distribution p^{eq}. While S may have transient peaks, the quantity H from Eq. 5.29 must be monotonically decreasing for all closed systems that obey a master equation with time independent transition amplitudes[129]. Explicit calculation of these quantities demonstrates that these requirements are satisfied, as shown in the attached document below, and serve to reinforce the physical plausibility of our model and its predictions.

Taken together, these observations form many of the major findings of the article included below[130], which concludes this text.

Adsorption-Desorption Kinetics of Soft Particles

Brendan Osberg, Johannes Nuebler, and Ulrich Gerland*
Theory of Complex Biosystems, Physik-Department, Technische Universität München, James-Franck-Strasse 1, 85748 Garching, Germany
(Received 27 April 2015; published 18 August 2015)

Adsorption-desorption processes are ubiquitous in physics, chemistry, and biology. Models usually assume hard particles, but within the realm of soft matter physics the adsorbing particles are compressible. A minimal 1D model reveals that softness fundamentally changes the kinetics: Below the desorption time scale, a logarithmic increase of the particle density replaces the usual Rényi jamming plateau, and the subsequent relaxation to equilibrium can be nonmonotonic and much faster than for hard particles. These effects will impact the kinetics of self-assembly and reaction-diffusion processes.

DOI: 10.1103/PhysRevLett.115.088301 PACS numbers: 82.70.Dd, 87.10.Mn, 87.15.R-

A broad range of physical, chemical, and biological systems feature adsorption processes in which particles are randomly deposited on an extended substrate [1]. Possible substrates include polymers [2], crystalline or amorphous surfaces [3], and membranes, while the particles can be small molecules, colloidal particles, macromolecules such as proteins [4,5], or even larger objects such as cells. In phenomenological models [6–8], the particles can also represent modified states of the substrate [6] or complexes formed with the substrate [8]. Generally, the filling of the substrate slows as the coverage increases, due to substrate saturation as well as jamming. Here, jamming refers to the process of reaching configurations where all gaps are smaller than the particles, prohibiting further filling. Jammed configurations constitute a nonthermal ensemble that has received considerable theoretical interest [9,10]. However, jamming is often only transient, as a small desorption rate allows the system to eventually reach thermal equilibrium in a nontrivial relaxation process during which the system loses memory of its jammed state [11–13].

Adsorption-desorption models usually assume hard particles that can neither be deformed nor overlap on the substrate. This assumption does not hold for systems such as soft colloidal particles [14], macromolecules like proteins [5], or complexes that can be forced into conformations with different effective sizes [15]. While the mechanical properties of soft particle systems are well studied [16], their adsorption-desorption kinetics have not been characterized. To explore this question, we analyze a minimal model with particles of finite stiffness ε binding to a one-dimensional (1D) substrate; see Fig. 1. Our model recovers the hard particle kinetics in the $\varepsilon \to \infty$ limit but demonstrates that the new parameter significantly enriches the kinetic behavior. Notably, softness can lead to nonmonotonic filling, where an initially empty substrate fills to a high "cramming density" ρ_{cr} before relaxing to the equilibrium density $\rho_{eq} < \rho_{cr}$. While this behavior may seem counterintuitive, it does not violate thermodynamic principles.

Our model is a generalization of the 1D car parking model [11,12] to soft particles. Models of this type are directly applicable to experimental systems with linear topology, and they also serve as a tractable theoretical framework to capture general kinetic phenomena [1]. For instance, 1D models show how the extremely slow relaxation of jammed systems [11,12] arises from the growing number of rearrangements required to make space for additional particles [13]. The same physics applies also in higher dimensions, e.g., to describe the slow densification of vibrated granular materials [13,17–19]. We discuss general implications of our 1D results further below.

Model.—Our model, illustrated in Fig. 1, describes the random adsorption and desorption of soft particles on a 1D substrate. The particles are assumed to have a finite interaction range set to 1, defining the length unit. Isolated particles adsorb at the bare rate r_+ per unit length and desorb with rate r_-. We are interested in the regime in which the rate ratio

$$r = r_+/r_- = e^{\mu} \tag{1}$$

is large, such that adsorption and desorption operate on very different time scales and a high equilibrium density is

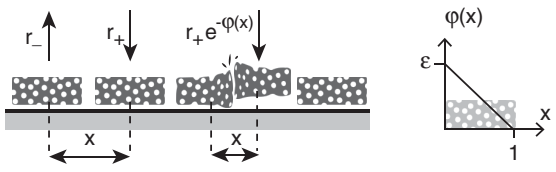


FIG. 1. One-dimensional adsorption-desorption model for soft particles. Overlapping adsorptions that require deformation are allowed but slowed down by the Boltzmann factor of the interaction energy $\varphi(x)$, which depends linearly on the center-to-center distance x. In the limit $\varepsilon \to \infty$ the hard particle model is recovered.

ultimately reached. For later convenience, Eq. (1) expresses r also in terms of the chemical potential of the noninteracting system, μ (throughout this Letter we set $k_BT = 1$ by choice of energy unit). A soft particle can attach even if it partially overlaps with its neighbor(s). We assume that overlaps are associated with an interaction $\varphi(x)$, where x denotes the center-to-center distance of adjacent particles. We primarily consider a potential that increases in proportion to proximity,

$$\varphi(x) = \begin{cases} \varepsilon(1-x) & \text{for } x \le 1, \\ 0 & \text{for } x > 1, \end{cases} \tag{2}$$

but also test to what extent the kinetic behavior depends on the shape of $\varphi(x)$. We note that potentials with finite stiffness ε can also serve as effective descriptions for hard particles that fluctuate between internal states with different effective lengths.

The interaction (2) modulates the reaction rates. An adsorption event alters the total interaction energy by $U(x_L, x_R) = \varphi(x_L) + \varphi(x_R) - \varphi(x_L + x_R)$, with the center-to-center distances to the left and right neighbors, x_L and x_R. Detailed balance requires that the modulated rates $\tilde{r}_+$, $\tilde{r}_-$ have the ratio $e^{\mu - U}$. As we do not seek to describe a specific experimental system, but rather to characterize generic effects of softness on adsorption-desorption kinetics, we can simply choose $\tilde{r}_- = r_-$ and

$$\tilde{r}_+(x_L, x_R) = r_+ e^{-U(x_L, x_R)}. \tag{3}$$

The Supplemental Material [20] discusses the effects of distributing the Boltzmann factor between adsorption and desorption.

Our model does not explicitly include lateral diffusion, although an effective form of lateral transport arises via desorption and readsorption [22]. We analyze the full stochastic kinetics of our model with simulations using the Gillespie method [23]. We also use a mean field description that characterizes the state of the system by the line density of particle spacings, $V(x,t)$, which obeys

$$\begin{aligned} \frac{\partial}{\partial t} V(x,t) = {} & 2\int_x^\infty dy V(y,t)\tilde{r}_+(x, y-x) - 2r_- V(x,t) \\ & - V(x,t)\int_0^x dy \tilde{r}_+(y, x-y) \\ & + r_- \int_0^x dy \frac{V(y,t)V(x-y,t)}{\rho(t)} \end{aligned} \tag{4}$$

and $\int_0^\infty dx\, xV(x,t) = 1$ (conservation of space). Since the number of voids equals the number of particles, the total particle density is $\rho(t) = \int_0^\infty dx\, V(x,t)$. Equation (4) describes the creation and destruction of voids of size x via adsorption within larger voids, desorption of a bounding particle, adsorption within the void, and the fusion of two smaller voids. In the last term of Eq. (4) the two-void density is approximated by the product of one-void densities, truncating the hierarchy of mean field equations at lowest order. Equation (4) recovers the mean field description of the car parking model [12] in the limit $\varepsilon \to \infty$. For $t \to \infty$, the equilibrium distribution $V_{\rm eq}(x) \propto e^{-\alpha x - \varphi(x)}$ is reached [20]. The equilibrium density can exceed unity since particles can overlap. See Ref. [20] for a comparison of the mean field and full model, the lattice equivalent of Eq. (4) used for all simulations, and a discussion of finite size effects.

Qualitative behavior.—Figure 2 characterizes the filling kinetics of an initially empty substrate. The time evolution of the total particle density $\rho(t)$ is shown in Fig. 2(a) for different stiffnesses ε, including the hard-core limit $\varepsilon \to \infty$. On the logarithmic time axis, the two time scales $1/r_+$ and $1/r_-$ roughly divide the kinetics into three separate stages. (i) Essentially unhindered adsorption for $t < 1/r_+$ with $\rho(t) \sim t$ independent of ε since interactions play only a minor role initially. At the end of this stage, most voids large enough for nonoverlapping adsorption are exhausted. (ii) For $1/r_+ < t < 1/r_-$, hard-core particles are in a jamming stage: Their density remains essentially constant at a plateau of $\rho_{\rm jam} \approx 0.748$, the Rényi limit [9]. In contrast, the density of soft particles keeps increasing, albeit only logarithmically, $\rho(t) - \rho_{\rm jam} \sim \log(t)$. (iii) In the third stage, $t > 1/r_-$, desorption becomes relevant and all systems relax to their equilibrium density $\rho_{\rm eq}$. However, the relaxation behavior changes dramatically with ε. Whereas the

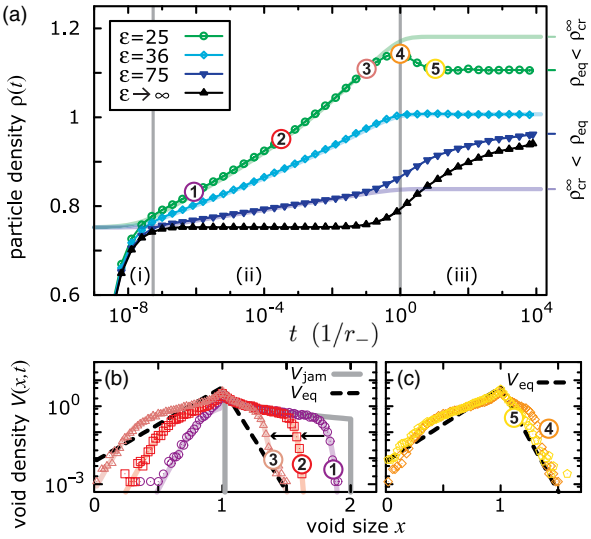


FIG. 2 (color online). Soft particle adsorption-desorption kinetics. (a) Density evolution for different stiffnesses ε (with $\mu = 20$). Symbols: stochastic simulations; transparent overlays: cramming dynamics of Eq. (6). Vertical lines indicate the adsorption and desorption time scales. (b),(c) Gap size distribution for $\varepsilon = 25$ at the time points marked in (a). At the onset of cramming, $V(x,t)$ is similar to the jammed distribution $V_{\rm jam}(x)$.

density of hard-core particles approaches ρ_{eq} extremely slowly from below, soft particles can either reach ρ_{eq} from below, arrive directly at ρ_{eq} at the onset of the third stage, or display a density overshoot before rapidly relaxing to ρ_{eq} from above.

The surprisingly rich kinetic behavior of Fig. 2(a) calls for a clarification of the underlying physics. How does the logarithmic behavior in stage (ii) arise from the softness and how generic is it? Under which conditions does nonmonotonic filling occur in an adsorption-desorption process that obeys detailed balance? Why is the relaxation to thermodynamic equilibrium much faster for soft particles? We address these questions in the remainder of this article by combining numerical analysis with analytical arguments.

Cramming.—The physics underlying the logarithmic regime (ii) of Fig. 2(a) is revealed by Fig. 2(b), which shows three consecutive snapshots of the void distribution $V(x,t)$. At the onset of stage (ii), $V(x,t)$ is similar to the known jammed distribution $V_{jam}(x)$ for irreversibly adsorbed hard particles [24],

$$V_{jam}(x) = 2\int_0^\infty dt\, t \exp\left[-(x-1)t - 2\int_0^t du \frac{1-e^{-u}}{u}\right] \tag{5}$$

[for $1 \le x < 2$, while $V_{jam}(x) = 0$ otherwise]: Both distributions display a dropoff for large gaps, while voids $x < 1$ are suppressed in $V(x,t)$ and are entirely forbidden in $V_{jam}(x)$. With increasing time, the dropoff in $V(x,t)$ progressively moves to smaller x as the largest available voids are filled (creating new voids with $x < 1$). This behavior indicates that the system's memory of the jammed configuration generated in stage (i) governs the "cramming" dynamics during stage (ii).

We quantify this physical picture by considering the Langmuir kinetics of reversibly filling the gaps in a jammed configuration. If $P(x,t)$ denotes the probability that a void of size x remains unfilled at time t, we have $P(x,t) = [\gamma_+(x)e^{[-(\gamma_+(x)+r_-)t]} + r_-]/[\gamma_+(x) + r_-]$, where $\gamma_+(x) = \int_0^x dy \tilde{r}_+(y, x-y)$ is an effective filling rate that combines all attachment possibilities. Neglecting multiple filling, this yields the cramming dynamics

$$V_{cr}(x,t) = P(x,t)V_{jam}(x) + 2\int_x^\infty dx'[1-P(x',t)] \times V_{jam}(x')\frac{\tilde{r}_+(x,x'-x)}{\gamma_+(x')}, \tag{6}$$

which can be considered an approximate solution to Eq. (4) for the cramming stage [20]. This distribution and the corresponding density $\rho_{cr}(t) = \int_0^\infty dx V_{cr}(x,t)$ are displayed as semitransparent lines in Fig. 2, showing that Eq. (6) captures the kinetics of stage (ii) very well. It also explains the logarithmic increase of $\rho(t)$: Approximating $\gamma_+(x)$ by its largest contribution, the drop in $V(x,t)$ moves as $\Delta x_{dr}(t) := 2 - x_{dr}(t) \approx \ln(r_+ t)/\varepsilon$ for our linear potential [20]. Given that the density is related to the area under the void size distribution, this yields [20]

$$\Delta\rho(t) \approx V_{jam}(2)\ln(r_+ t)/\varepsilon, \tag{7}$$

which rationalizes the logarithmic time dependence and predicts how the dynamics slow down with increasing stiffness ε.

Nonmonotonic density.—Remarkably, $\rho(t)$ can transiently exceed the equilibrium density. We now show that this is a result of desorption erasing the memory of stage (i) that was preserved during stage (ii). We first note that the density $\rho_{cr}(t)$ obtained above (by assuming that only cramming is reversible while the underlying jammed configuration is preserved) saturates towards a value ρ_{cr}^∞, which can be smaller than the equilibrium density ρ_{eq} [the blue line in Fig. 2(a)], or can exceed it (the green line). Figure 3(a) shows that the parameter regime where $\rho_{cr}^\infty > \rho_{eq}$ is virtually identical with the regime where the maximal density ρ_{max} exceeds ρ_{eq}. Thus, a nonmonotonic density $\rho(t)$ occurs whenever suppressing rearrangements leads to a larger equilibrium density than allowing for them.

To elucidate the minimal requirements for nonmonotonic filling and to study the phenomenon from the perspective of nonequilibrium thermodynamics, it is useful to consider the case of dimers on a discrete lattice. For soft dimers, which can overlap by a single site at the energetic cost $\varepsilon/2$, nonmonotonic filling occurs in a similar parameter range as in the continuum model; see Fig. 3(b). In fact, the boundaries in (ε, μ) space can be understood with a simple argument: A density overshoot is possible only if single overlaps occur faster than desorption and simultaneous overlaps with both neighboring particles are rare, which translates into the condition [20]

$$\varepsilon/2 < \mu < \varepsilon. \tag{8}$$

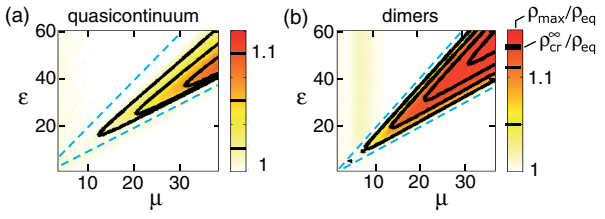


FIG. 3 (color online). Phase diagram in (ε, μ) space for the density overshoot with (a) quasicontinuum particles and (b) soft dimers. The overshoot is the ratio of the maximum and the equilibrium densities ρ_{max}/ρ_{eq} observed numerically (color coded). Overlayed are contour lines of $\rho_{cr}^\infty/\rho_{eq}$, where ρ_{cr}^∞ is the asymptotic cramming density derived from Eq. (6). Dashed lines indicate the regime of Eq. (8).

This regime is indicated by the dashed lines in Fig. 3. Interestingly, this regime also encompasses the nonmonotonic region of the continuum model, showing that Eq. (8) provides a necessary (but not sufficient) condition for the class of models that we consider.

Why can a density overshoot occur for soft but not for hard particles? One way to address this question is to consider the large-ε behavior of the lattice equivalent of Eq. (4). After the initial filling stage, the only relevant components of the void density vector for soft dimers are the ones for overlapping dimers, for adjacent dimers, and for single empty sites. Space conservation then reduces the dynamics to a two-dimensional first-order ordinary differential equation, which permits nonmonotonic behavior (see the Supplemental Material [20] for details). However, in the limit $\varepsilon \to \infty$, the dynamics become effectively one dimensional, and hence monotonic, since the overlapping dimer degree of freedom is lost. Thus, the mathematical mechanism within the mean field description is dimensional reduction.

Nonequilibrium thermodynamics.—Figure 4 shows the filling dynamics $\rho(t)$ of dimers together with two thermodynamic quantities, S and H. The time-dependent entropy $S(t) = -(1/L)\sum_n p_n(t) \log p_n(t)$ can be computed via the occupation probabilities $p_n(t)$ of all configurations of the system (we measure the system size L in units of the particle size). It displays a nonmonotonic behavior, not only for soft dimers (the dashed line) but also for hard dimers (the solid line), which is not surprising since our system is initially far from equilibrium. That $S(t)$ first rises and later decreases during relaxation is consistent with the evolution from a single initial state (empty) through disordered intermediate states to a highly ordered equilibrium state (relevant stages are sketched in Fig. 4). However, the relative entropy $H(t) = (1/L)\sum_n p_n(t) \log[p_n(t)/p_n^{\mathrm{eq}}]$ (with respect to the equilibrium state) must monotonically decrease for a system described by a discrete master equation that obeys detailed balance [25]. Figure 4 shows that the density overshoot is compatible with this fundamental theorem.

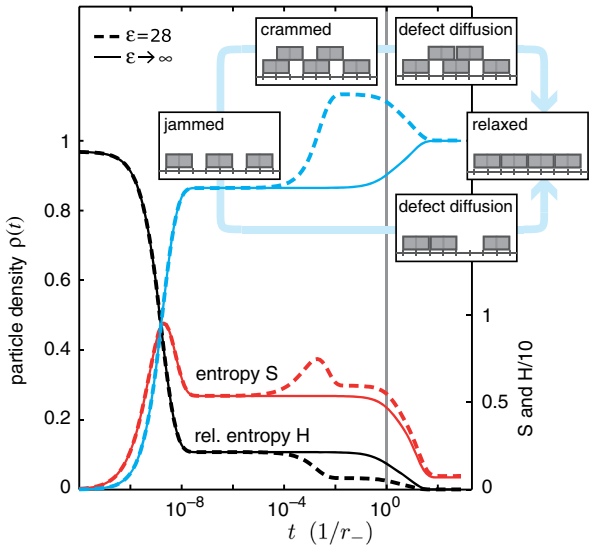


FIG. 4 (color online). Evolution of the density (left axis) and thermodynamic quantities (right axis) for hard (solid) and soft (dashed) dimers ($\mu = 20$). Soft dimers may overlap by one site at the cost of the interaction energy $\varepsilon/2 = 14$. Relevant stages for hard (bottom path) and soft (top path) dimers are sketched.

Relaxation kinetics.—Figure 2 indicated that relaxation to equilibrium is faster for soft particles than for hard. To clarify whether the relaxation behavior is qualitatively different, we examine how $|\rho(t) - \rho_{\mathrm{eq}}|$ approaches zero in Fig. 5 (and in more detail in the Supplemental Material [20]). For dimers, the relaxation behavior is actually the same for hard and soft particles, due to a particle-hole symmetry [20]: Both jamming (for hard particles) and cramming (for soft particles) lead to configurations with "defects," which undergo a diffusion-annihilation process. For hard particles, the defects are isolated, unoccupied lattice sites, and diffusion occurs via desorption of an adjacent particle followed by immediate adsorption of a particle into the gap [22]. The progressive dilution of defects leads to the power law behavior $|\rho(t) - \rho_{\mathrm{eq}}| \sim t^{-1/2}$, which holds until the finite defect creation rate balances the diffusion-annihilation process. For soft particles, the defects are sites with double occupancy, which by a similar mechanism lead to the same power law behavior and even a "mirror symmetry" of the relaxation curve around density one [20].

This symmetry is broken when the particle size is increased to k-mers with $k > 2$ (since reactions then occur between defects of different sizes). Figure 5 shows that for hard particles the relaxation behavior becomes *slower* as k increases, while it becomes *faster* for soft particles. The scaling approaches logarithmic behavior in the limit $k \to \infty$ for hard particles [11,12], while it approaches exponential behavior for soft particles. Qualitatively, this is explained by the fact that the soft interaction "guides" attaching particles to the most favorable positions (reducing the entropic barrier for the rearrangements required for equilibration). Other shapes of the interaction potential lead

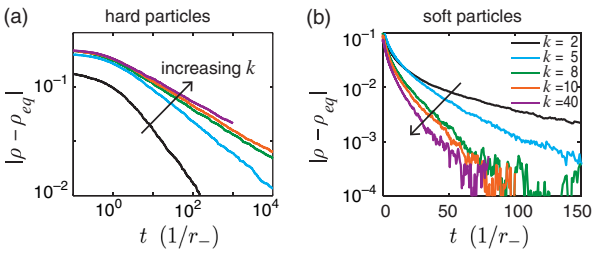


FIG. 5 (color online). Equilibration behavior of hard (a) and soft (b) particles with different sizes k on a lattice ($\mu = 30$). For hard particles, the relaxation kinetics slows with increasing k, while it becomes faster for soft particles ($\varepsilon = 36$).

to the same behavior as long as the repulsion is sufficiently soft [20].

Discussion.—We have shown that the adsorption-desorption kinetics of soft particles differs fundamentally from that of hard particles in at least three aspects: (i) the jamming behavior, with a gradual density increase instead of the Rényi plateau, (ii) a density overshoot, which can occur only for soft particles, on a time scale set by the desorption rate, and (iii) the relaxation behavior, which for soft particles becomes faster with increasing particle size (on a lattice), while hard particles show the opposite trend. We performed our analysis for a minimal model and showed that our qualitative conclusions are not sensitive to details such as the precise shape of the repulsive interaction potential and the way in which it affects the kinetic rates.

Models within this class are directly relevant in biophysics, for instance, in describing the binding of dimeric kinesins to microtubules [26] (where the softness stems from the ability to bind with either one or two head domains) or the assembly of nucleosome arrays [15] (where the softness arises from transient, thermally induced DNA unwrapping). While our analysis was limited to 1D substrates, we expect that much of the qualitative phenomenology carries over to 2D substrates. An interesting 2D experimental system is protein adsorption from blood plasma, which can show nonmonotonic surface density [27–29]. This effect is not well understood but is usually interpreted in a two-species scenario where a fast-binding protein is replaced by a slower but stronger-binding competitor. Our findings suggest that even a single soft protein species could generate nonmonotonic densities.

Another field of application is the physics of vibrated granular materials. The sluggish kinetics of these systems is nonadiabatic and shows signs of broken ergodicity [30,31]. One-dimensional models have already been useful to describe certain aspects of these kinetics phenomenologically [13,32]. The introduction of an effective soft-core interaction will provide a valuable new dimension in the parameter space of such phenomenological descriptions. More generally, it will be interesting to explore how the rich adsorption-desorption kinetics of soft particles couples to other kinetic processes. For instance, it should modify the collective dynamics of molecular motors, reaction-diffusion processes that involve a lower-dimensional substrate, and substrate-guided self-assembly processes.

*gerland@tum.de

[1] W. Evans, Rev. Mod. Phys. **65**, 1281 (1993).
[2] J. D. McGhee and P. H. von Hippel, J. Mol. Biol. **86**, 469 (1974).
[3] J. Krug and T. Michely, *Islands, Mounds and Atoms*, Springer Series in Surface Sciences (Springer, New York, 2004).
[4] J. Feder, J. Theor. Biol. **87**, 237 (1980).
[5] J. Talbot, G. Tarjus, P. R. Van Tassel, and P. Viot, Colloids Surf., A **165**, 287 (2000).
[6] J. Flory, J. Am. Chem. Soc. **61**, 1518 (1939).
[7] J. Toner and G. Y. Onoda, Phys. Rev. Lett. **69**, 1481 (1992).
[8] R. Padinhateeri and J. F. Marko, Proc. Natl. Acad. Sci. U.S.A. **108**, 7799 (2011).
[9] A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **3**, 109 (1958); Selected translations in mathematical statistics and probability **4**, 203 (1963).
[10] S. F. Edwards and R. B. S. Oakeshott, Physica (Amsterdam) **157A**, 1080 (1989).
[11] X. Jin, G. Tarjus, and J. Talbot, J. Phys. A **27**, L195 (1994).
[12] P. L. Krapivsky and E. Ben-Naim, J. Chem. Phys. **100**, 6778 (1994).
[13] E. Ben-Naim, J. B. Knight, E. R. Nowak, H. M. Jaeger, and S. R. Nagel, Physica (Amsterdam) **123D**, 380 (1998).
[14] J. Mattsson, H. M. Wyss, A. Fernandez-Nieves, K. Miyazaki, Z. Hu, D. R. Reichman, and D. A. Weitz, Nature (London) **462**, 83 (2009).
[15] B. Osberg, J. Nuebler, P. Korber, and U. Gerland, Nucleic Acids Res. **42**, 13633 (2014).
[16] M. van Hecke, J. Phys. Condens. Matter **22**, 033101 (2010).
[17] J. B. Knight, C. G. Fandrich, C. N. Lau, H. M. Jaeger, and S. R. Nagel, Phys. Rev. E **51**, 3957 (1995).
[18] J. Talbot, G. Tarjus, and P. Viot, Phys. Rev. E **61**, 5429 (2000).
[19] P. Richard, M. Nicodemi, R. Delannay, P. Ribière, and D. Bideau, Nat. Mater. **4**, 121 (2005).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.088301, which includes Ref. [21], for elaborated presentations of several arguments and numerical methods.
[21] W. Möbius, B. Osberg, A. M. Tsankov, O. J. Rando, and U. Gerland, Proc. Natl. Acad. Sci. U.S.A. **110**, 5719 (2013).
[22] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, England, 2010).
[23] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[24] G. Tarjus and P. Viot, Phys. Rev. E **69**, 011307 (2004).
[25] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Third ed. (Elsevier, New York, 2007).
[26] E. Frey and A. Vilfan, Chem. Phys. **284**, 287 (2002).
[27] S. L. Hirsh, D. R. McKenzie, N. J. Nosworthy, J. A. Denman, O. U. Sezerman, and M. M. M. Bilek, Colloids Surf., B **103**, 395 (2013).
[28] L. Vroman, A. L. Adams, G. C. Fischer, and P. C. Munoz, Blood **55**, 156 (1980).
[29] J. L. Brash and P. ten Hove, Thromb. Haemostas. **51**, 326 (1984).
[30] A. J. Kolan, E. R. Nowak, and A. V. Tkachenko, Phys. Rev. E **59**, 3094 (1999).
[31] F. Paillusson and D. Frenkel, Phys. Rev. Lett. **109**, 208001 (2012).
[32] E. R. Nowak, J. B. Knight, E. Ben-Naim, H. M. Jaeger, and S. R. Nagel, Phys. Rev. E **57**, 1971 (1998).

A.1 Two Particle Correlation in the Large L Limit

In Chapter 2, we reduced the two-particle correlation function to

$$\rho(x|0) = \frac{\mathrm{Tr}\left\{F_s T^x F_s T^{L-x}\right\}}{\mathrm{Tr}\left\{F_s T^L\right\}}, \tag{A.1}$$

where the numerator was rewritten as $\mathrm{Tr}\left\{Fe \cdot D^x \cdot Fe \cdot D^{L-x}\right\}$. For notational simplicity we write $\tilde{L} = L - x$ and see that

$$Fe \cdot D^{\tilde{L}} = \begin{bmatrix} Fe_{00} & Fe_{01} & Fe_{02} & \\ Fe_{10} & Fe_{11} & Fe_{12} & \\ & & & \ddots \end{bmatrix} \cdot \begin{bmatrix} \lambda_0^{\tilde{L}} & 0 & 0 & \\ 0 & \lambda_1^{\tilde{L}} & 0 & \\ 0 & 0 & \lambda_2^{\tilde{L}} & \\ & & & \ddots \end{bmatrix} \tag{A.2}$$

$$= \begin{bmatrix} Fe_{00}\lambda_0^{\tilde{L}} & Fe_{01}\lambda_1^{\tilde{L}} & Fe_{02}\lambda_2^{\tilde{L}} & \\ Fe_{10}\lambda_0^{\tilde{L}} & Fe_{11}\lambda_1^{\tilde{L}} & Fe_{12}\lambda_2^{\tilde{L}} & \\ Fe_{20}\lambda_0^{\tilde{L}} & Fe_{21}\lambda_1^{\tilde{L}} & Fe_{22}\lambda_2^{\tilde{L}} & \\ & & & \ddots \end{bmatrix}. \tag{A.3}$$

Proceeding further,

$$D^x \cdot Fe \cdot D^{\tilde{L}} = \begin{bmatrix} \lambda_0^x & 0 & 0 & \\ 0 & \lambda_1^x & 0 & \\ 0 & 0 & \lambda_2^x & \\ & & & \ddots \end{bmatrix} \cdot \begin{bmatrix} Fe_{00}\lambda_0^{\tilde{L}} & Fe_{01}\lambda_1^{\tilde{L}} & Fe_{02}\lambda_2^{\tilde{L}} & \\ Fe_{10}\lambda_0^{\tilde{L}} & Fe_{11}\lambda_1^{\tilde{L}} & Fe_{12}\lambda_2^{\tilde{L}} & \\ Fe_{20}\lambda_0^{\tilde{L}} & Fe_{21}\lambda_1^{\tilde{L}} & Fe_{22}\lambda_2^{\tilde{L}} & \\ & & & \ddots \end{bmatrix} \tag{A.4}$$

$$= \begin{bmatrix} Fe_{00}\lambda_0^{x+\tilde{L}} & Fe_{01}\lambda_0^x\lambda_1^{\tilde{L}} & Fe_{02}\lambda_0^x\lambda_2^{\tilde{L}} & \\ Fe_{10}\lambda_1^x\lambda_0^{\tilde{L}} & Fe_{11}\lambda_1^{x+\tilde{L}} & Fe_{12}\lambda_1^x\lambda_2^{\tilde{L}} & \\ Fe_{20}\lambda_2^x\lambda_0^{\tilde{L}} & Fe_{21}\lambda_2^x\lambda_1^{\tilde{L}} & Fe_{22}\lambda_2^{x+\tilde{L}} & \\ & & & \ddots \end{bmatrix}. \tag{A.5}$$

One further matrix multiplication by Fe produces diagonal elements of $Fe \cdot D^x \cdot Fe \cdot D^{\tilde{L}}$ given by:

$$\begin{bmatrix} \lambda_0^{\tilde{L}} \sum_{\nu=0} Fe_{0\nu}Fe_{\nu 0}\lambda_\nu^x & & & \\ & \lambda_1^{\tilde{L}} \sum_{\nu=0} Fe_{1\nu}Fe_{\nu 1}\lambda_\nu^x & & \\ & & \lambda_2^{\tilde{L}} \sum_{\nu=0} Fe_{2\nu}Fe_{\nu 2}\lambda_\nu^x & \\ & & & \ddots \end{bmatrix}. \tag{A.6}$$

The finite non-diagonal elements from Eq. A.6 can be ignored when taking the trace:

$$\mathrm{Tr}\left\{Fe\cdot D^x\cdot Fe\cdot D^{\tilde{L}}\right\}=\lambda_0^{\tilde{L}}\sum_{\nu=0}Fe_{0\nu}Fe_{\nu 0}\lambda_\nu^x+\lambda_1^{\tilde{L}}\sum_{\nu=0}Fe_{1\nu}Fe_{\nu 1}\lambda_\nu^x+\ldots \tag{A.7}$$

Recalling from Chapter 2 we determined $\mathrm{Tr}\left\{FsT^L\right\}=\sum_{\nu=0}\left(Fe_{\nu\nu}\lambda_\nu^L\right)$, the conditional probability from the beginning of this section becomes

$$\begin{aligned}\rho(x|0) &= \frac{\mathrm{Tr}\left\{F_sT^xF_sT^{L-x}\right\}}{\mathrm{Tr}\left\{F_sT^L\right\}}, &\text{(A.8)}\\ &= \frac{\lambda_0^{\tilde{L}}\sum_{\nu=0}Fe_{0\nu}Fe_{\nu 0}\lambda_\nu^x+\lambda_1^{\tilde{L}}\sum_{\nu=0}Fe_{1\nu}Fe_{\nu 1}\lambda_\nu^x+\ldots}{\sum_{\nu=0}\left(Fe_{\nu\nu}\lambda_\nu^L\right)} &\text{(A.9)}\end{aligned}$$

and upon taking the limit $\tilde{L}\to\infty$, the largest eigenvalue of this series dominates:

$$\rho(x|0)=\frac{1}{Fe_{00}}\sum_{\nu=0}Fe_{0\nu}Fe_{\nu 0}\left(\frac{\lambda_\nu}{\lambda_0}\right)^x, \tag{A.10}$$

as given in Eq. 2.15

A.2 The Amoeba Method

In Chapter 4, we discussed a numeric scheme to infer the potential landscape from the equilibrium density of particles along a 1-D substrate of length L for a given interaction potential between neighboring particles. The basis of this solution depended on nominating candidate landscape potentials V_i with which to predict a density pattern that best reproduces experimental data –or, more precisely, one which minimizes the sum of squared error f with said data.

How the initial guesses for V_i are nominated is inconsequential, provided they are linearly independent, and the vectors between them span the space. In other words, for each initial guesses V_i there must exist a set of coefficients c_j such that all points in L space are expressible as $V_i + \sum_j c_j(V_j - V_i)$. Section 4.3.1 provides one means of doing this, although this is by no means unique.

Once initial values are chosen, they are ranked by their f_i values, and the poorest candidate with f_h (h denotes a 'high' functional value) are transformed in search of better positions. To perform these transformations, the centroid of the points $\overline{V}$, or mean positioned is defined as a reference, about which f_h is reflection, expanded or all points are contracted, until convergence at an optimum solution. Fig. A.1 shows the operations that are performed on the set of V_i's and is based on the same flow diagram in Ref. [92]. Determination of this solution used standard values[92, 91] for the reflection coefficient $\alpha = 1$, expansion coefficient $\gamma = 2$, and contraction coefficient $\beta = \frac{1}{2}$. In general, the tolerance of position coordinates, $PTOL$ is defined by the precision with which the position is reported, although this is rarely the limiting convergence criteria. Convergence of f values within $FTOL = 10^{-4}$, or one part in ten thousand of the smallest f value typically defined the exit criteria of the algorithm. Upon convergence, the resulting point was then used as an initial guess V_i, and the whole process was started again. This was repeated until successive iterations were, themselves, converged within $FTOL$, and the final result was taken.

To test the reliability of this method, a known potential was used to generate a density profile to see whether the algorithm was able to infer that potential from the density and neighbor interaction alone. Fig. A.2 shows the results of this proof of concept for a test system of size $L = 100$, a=10 with various arbitrary features embedded in the landscape, in addition to several intermediate candidate landscapes that were generated in the process of converging to the optimum solution. Starting from a flat landscape guess (green), proceeding through intermediate guesses (blue) the output of the algorithm (red) is consistent with the exact solution (black). With this proof of concept in mind, the algorithm was then applied to real data as described in Chapter 4. Although Fig. A.2 represents only one regime of possible neighbor interactions (where positioning is rather Boltzmann distributed), similar tests in other regimes show the robustness of the method to other interaction strengths and at higher densities.

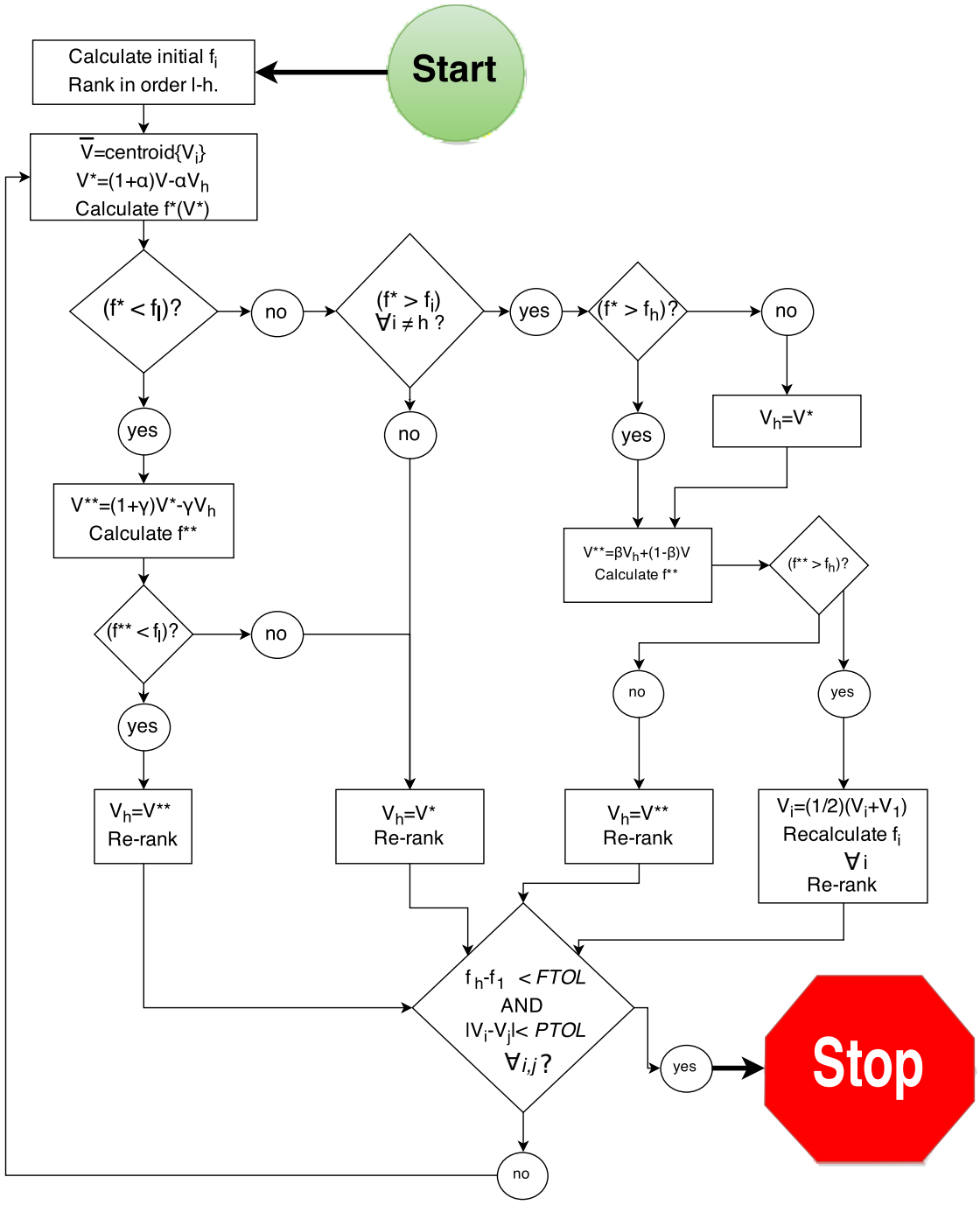


Figure A.1: The alogorithm by which the amoeba method converges at an optimal solution to the minimization of a multivariable function. Boolean conditionals, in diamonds, are followed by ovals for "yes"(true) or "no"(false) leading to operations in rectangles where equality denotes assignment. The reflection, expansion, and contraction transformations in these rectangles are iterated until convergence is reached. Adapted from Ref. [92].

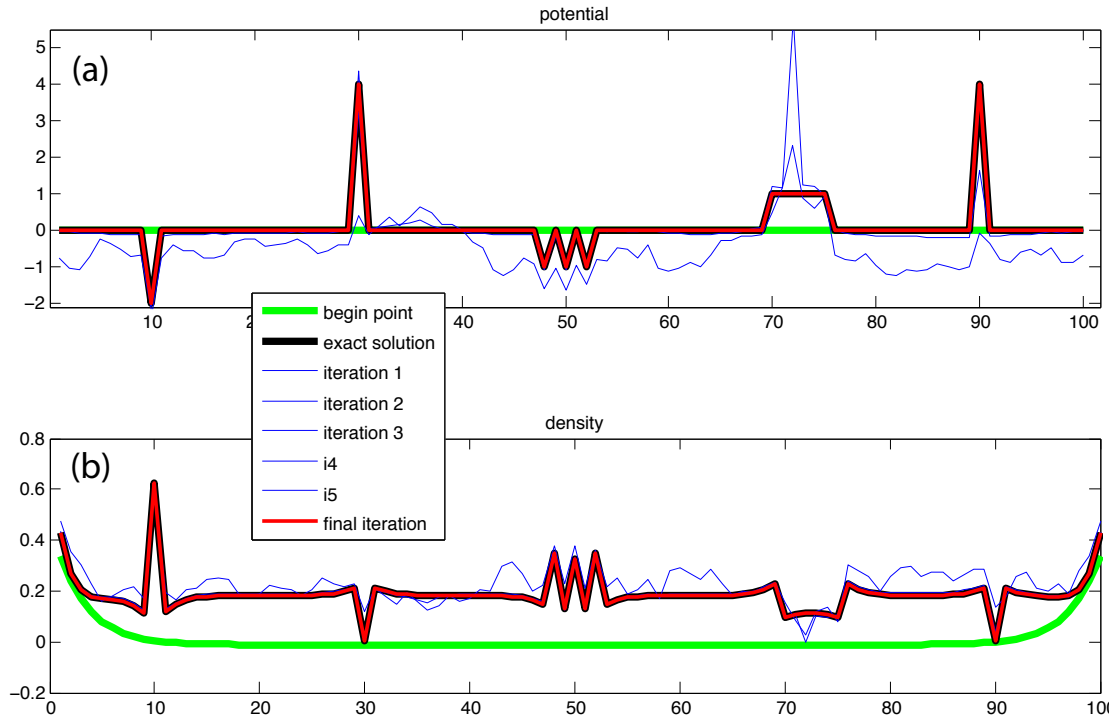


Figure A.2: (a) Various nucleosome adsorption landscapes $V(x)$ including a flat initial guess (green), several intermediate profiles (blue), and a converged output (red) that is highly consistent with the exact solution (black). (b) particle density profiles calculated from the above potentials with corresponding color codes. Again, the final projected density in red, is plotted on top of the exact density in black.

A.3 Acknowledgements

I would like to thank CeNS, the Elite Network of Bavaria, and Biosysnet, for providing the opportunity to work and live in Munich during this time. I would like to thank my supervisor Ulrich Gerland, for all of the advice, guidance, and patient support that were necessary throughout this program, and to Johannes Nuebler for the helpful collaboration. Thanks for Kien Nguyen and Jorge Trojanoski for their diligent hard work and contributions. Thanks to Philipp Korber, Nils Krietenstein, Maria Walker, and group for making available both the biological data from their work, as well as their wealth of knowledge on a wide array of subject matter that has proven indispensable throughout the process of this work.

At various stages, content and attached published works underwent proofreading for which I thank Mary Clancy, Caroline Dinsel, Patrick Hillenbrand, Florian Hinzpeter, Tiago Ramalho, Emmanuel Reithmann, Filipe Tostevin, Markus Weber, and Michael Wolff.

On a more personal level, thanks to Mirjam Burkhardt for help in getting through some of the rough patches. Special thanks to my father, Lars Osberg, for inspiring the pursuit of science, and for help and advice in managing some of the non-scientific challenges of academic life. Thanks as well to my siblings Natasha Scott and Spencer Osberg for cheering from afar, and providing the necessary support and encouragement for getting across the finish line.

Bibliography

[1] Roger D. Kornberg. Chromatin Structure : A Repeating Unit of Histones and DNA; Chromatin structure is based on a repeating unit of eight histone molecules and about 200 DNA base pairs. *Science*, 184(4139):868–871, 1974.

[2] J. T. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 73(6):1897–1901, 1976.

[3] Margot P. Scheffer, Mikhail Eltsov, and Achilleas S. Frangakis. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 108(41):16992–7, 2011.

[4] Yoshinori Nishino, Mikhail Eltsov, Yasumasa Joti, Kazuki Ito, Hideaki Takata, Yukio Takahashi, Saera Hihara, Achilleas S. Frangakis, Naoko Imamoto, Tetsuya Ishikawa, and Kazuhiro Maeshima. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal*, 31(7):1644–1653, 2012.

[5] Howard Cedar and Yehudit Bergman. Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews. Genetics*, 10(5):295–304, 2009.

[6] Oliver Bell, Vijay K. Tiwari, Nicolas H. Thomä, and Dirk Schübeler. Determinants and dynamics of genome accessibility. *Nature reviews. Genetics*, 12(8):554–64, 2011.

[7] Ian B. Dodd, Mille A. Micheelsen, Kim Sneppen, and Geneviève Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129(4):813–22, 2007.

[8] Karthika Raghavan, Heather J. Ruskin, Dimitri Perrin, Francois Goasmat, and John Burns. Computational Micromodel for Epigenetic Mechanisms. *PLoS ONE*, 5(11):e14031, 2010.

[9] Julien Riposo and Julien Mozziconacci. Nucleosome positioning and nucleosome stacking: two faces of the same coin. *Molecular BioSystems*, 8(4):1172–8, 2012.

[10] Roger D. Kornberg. The location of nucleosomes in chromatin: specific or statistical? *Nature*, 292(13):579–580, 1981.

[11] Roger D. Kornberg and Lubert Stryer. Statistical distributions of nucleosomes : nonrandom locations by a stochastic mechanism. *Nucleic Acids Research*, 16(14):6677–6690, 1988.

[12] Tom Chou. An exact theory of histone-DNA adsorption and wrapping. *Europhysics Letters (EPL)*, 62(5):753–759, 2003.

[13] Anthony T. Annunziato. Assembling chromatin: The long and winding road. *Biochimica et Biophysica Acta*, 1819(3-4):196–210, 2012.

[14] Eran Segal and Jonathan Widom. What controls nucleosome positions? *Trends in genetics*, 25(8):335–43, 2009.

[15] Travis N. Mavrich, Ilya P. Ioshikhes, Bryan J. Venters, Cizhong Jiang, Lynn P. Tomsho, Ji Qi, Stephan C. Schuster, Istvan Albert, and B. Franklin Pugh. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research*, 18(7):1073–83, 2008.

[16] Zhenhai Zhang, Christian J. Wippo, Megha Wal, Elissa Ward, Philipp Korber, and B. Franklin Pugh. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, 332(6032):977–80, 2011.

[17] Assaf Weiner, Amanda Hughes, Moran Yassour, Oliver J. Rando, and Nir Friedman. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome research*, 20(1):90–100, 2010.

[18] Barbara Celona, Assaf Weiner, Francesca Di Felice, Francesco M. Mancuso, Elisa Cesarini, Riccardo L. Rossi, Lorna Gregory, Dilair Baban, Grazisa Rossetti, Paolo Grianti, Massimiliano Pagani, Tiziana Bonaldi, Jiannis Ragoussis, Nir Friedman, Giorgio Camilloni, Marco E. Bianchi, and Alessandra Agresti. Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biology*, 9(6):e1001086, 2011.

[19] Kristin Brogaard, Liqun Xi, Ji-Ping Wang, and Jonathan Widom. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404):496–501, 2012.

[20] Lu Bai and Alexandre V. Morozov. Gene regulation by nucleosome positioning. *Trends in Genetics*, 26(11):476–83, 2010.

[21] J. G. Henikoff, J. A. Belsky, K. Krassovsky, D. M. MacAlpine, and S. Henikoff. Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45):18318–18323, 2011.

[22] Cizhong Jiang and B. Franklin Pugh. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews. Genetics*, 10(3):161–172, 2009.

[23] Marta Radman-Livaja and Oliver J. Rando. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology*, 339(2):258–266, 2010.

[24] Oliver J. Rando and Howard Y. Chang. Genome-wide views of chromatin structure. *Annual review of biochemistry*, 78:245–271, 2009.

[25] Cedric R. Clapier and Bradley R. Cairns. The biology of chromatin remodeling complexes. *Annual review of biochemistry*, 78:273–304, 2009.

[26] Ronen Sadeh and C. David Allis. Genome-wide Re-Modeling of Nucleosome Positions. *Cell*, 147(2):263–6, 2011.

[27] Harm van Bakel, Kyle Tsui, Marinella Gebbia, Sanie Mnaimneh, Timothy R. Hughes, and Corey Nislow. A Compendium of Nucleosome and Transcript Profiles Reveals Determinants of Chromatin Architecture and Transcription. *PLoS Genetics*, 9(5):e1003479, 2013.

[28] Judith B. Zaugg and Nicholas M. Luscombe. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome research*, 22(1):84–94, 2012.

[29] Hinrich Boeger, Joachim Griesenbeck, and Roger D. Kornberg. Nucleosome Retention and the Stochastic Nature of Promoter Chromatin Remodeling for Transcription. *Cell*, 133(4):716–726, 2008.

[30] G. Chevereau, L. Palmeira, C. Thermes, A. Arneodo, and C. Vaillant. Thermodynamics of intragenic nucleosome ordering. *Physical Review Letters*, 103(18):188103, 2009.

[31] Sushma Shivaswamy, Akshay Bhinge, Yongjun Zhao, Steven Jones, Martin Hirst, and Vishwanath R. Iyer. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biology*, 6(3):e65, 2008.

[32] Wolfram Möbius and Ulrich Gerland. Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS computational biology*, 6(8):e1000891, 2010.

[33] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fur Physik*, 31(1):253–258, 1925.

[34] Stephen G. Brush. History of the Lenz-Ising Model. *Reviews of Modern Physics*, 39(4):883–893, 1967.

[35] Vladimir B Teif. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to O_R operator of phage λ. *Nucleic Acids Research*, 35(11):e80, 2007.

[36] Wolfram Möbius, Brendan Osberg, Alexander M. Tsankov, Oliver J. Rando, and Ulrich Gerland. Toward a unified physical model of nucleosome patterns flanking transcription start sites. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14):5719–24, 2013.

[37] Yi-der Chen. Binding of n-mers to one-dimensional lattices with longer than close-contact interactions. *Biophysical Chemistry*, 27(1):59–65, 1987.

[38] Zevi W. Salsburg, Robert W. Zwanzig, and John G. Kirkwood. Molecular Distribution Functions in a One-Dimensional Fluid. *The Journal of Chemical Physics*, 21(6):1098–1107, 1953.

[39] Alberto Robledo. The liquidsolid transition of the hard sphere system from uniformity of the chemical potential. *The Journal of Chemical Physics*, 72(3):1701, 1980.

[40] J. D. Anderson, A. Thåström, and J. Widom. Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. *Molecular and cellular biology*, 22(20):7147–7157, 2002.

[41] W. J. Koopmans, R. Buning, T. Schmidt, and J. van Noort. spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophysical journal*, 97(1):195–204, 2009.

[42] Gu Li and Jonathan Widom. Nucleosomes facilitate their own invasion. *Nature structural & molecular biology*, 11(8):763–9, 2004.

[43] Maik Engeholm, Martijn de Jager, Andrew Flaus, Ruth Brenk, John van Noort, and Tom Owen-Hughes. Nucleosomes can invade DNA territories occupied by their neighbors. *Nature structural & molecular biology*, 16(2):151–8, 2009.

[44] J. D. Anderson and J. Widom. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology*, 296(4):979–87, 2000.

[45] Hannah S. Tims, Kaushik Gurunathan, Marcia Levitus, and Jonathan Widom. Dynamics of nucleosome invasion by DNA binding proteins. *Journal of Molecular Biology*, 411(2):430–448, 2011.

[46] Maria Hondele, Tobias Stuwe, Markus Hassler, Felix Halbach, Andrew Bowman, Elisa T. Zhang, Bianca Nijmeijer, Christiane Kotthoff, Vladimir Rybin, Stefan Amlacher, Ed Hurt, and Andreas G Ladurner. Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature*, 499(7456):111–4, 2013.

[47] Gu Li, Marcia Levitus, Carlos Bustamante, and Jonathan Widom. Rapid spontaneous accessibility of nucleosomal DNA. *Nature structural & molecular biology*, 12(1):46–53, 2005.

[48] Vladimir B. Teif, Ramona Ettig, and Karsten Rippe. A lattice model for transcription factor access to nucleosomal DNA. *Biophysical Journal*, 99(8):2597–607, 2010.

[49] Răzvan V. Chereji and Alexandre V Morozov. Ubiquitous nucleosome crowding in the yeast genome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14):5236–41, 2014.

[50] Ji-Ping Wang, Yvonne Fondufe-Mittendorf, Liqun Xi, Guei-Feng Tsai, Eran Segal, and Jonathan Widom. Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae. *PLoS Computational Biology*, 4(9):e1000175, 2008.

[51] Daniel T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[52] Kazuhiro Yamada, Timothy D. Frouws, Brigitte Angst, Daniel J. Fitzgerald, Carl DeLuca, Kyoko Schimmele, David F. Sargent, and Timothy J. Richmond. Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature*, 472(7344):448–53, 2011.

[53] Lisa R. Racki, Janet G. Yang, Nariman Naber, Peretz D. Partensky, Ashley Acevedo, Thomas J. Purcell, Roger Cooke, Yifan Cheng, and Geeta J. Narlikar. The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature*, 462(7276):1016–21, 2009.

[54] Evgeniy A. Ozonov and Erik van Nimwegen. Nucleosome Free Regions in Yeast Promoters Result from Competitive Binding of Transcription Factors That Interact with Chromatin Modifiers. *PLoS Computational Biology*, 9(8):e1003181, 2013.

[55] Geeta J. Narlikar, Hua Ying Fan, and Robert E. Kingston. Cooperation between complexes that regulate chromatin structure and transcription. *Cell*, 108(4):475–487, 2002.

[56] Walter J. Jessen, Scott Hoose, Jessica Kilgore, and Michael P. Kladde. Active PHO5 chromatin encompasses variable numbers of nucleosomes at individual promoters. *Nature structural & molecular biology*, 13(3):256–263, 2006.

[57] Itay Tirosh and Naama Barkai. Two strategies for gene regulation by promoter nucleosomes. *Genome research*, 18(7):1084–91, 2008.

[58] Ionas Erb and Erik van Nimwegen. Transcription factor binding site positioning in yeast: Proximal promoter motifs characterize Tata-Less promoters. *PLoS ONE*, 6(9):e24279, 2011.

[59] Karolin Luger. Dynamic nucleosomes. *Chromosome Research*, 14(1):5–16, 2006.

[60] Dana. J. Huebert, Pei-Fen. Kuan, Sündüz. Keles, and Audrey. P. Gasch. Dynamic Changes in Nucleosome Occupancy Are Not Predictive of Gene Expression Dynamics but Are Linked to Transcription and Chromatin Regulators. *Molecular and Cellular Biology*, 32(9):1645–53, 2012.

[61] Rutger Hermsen, Sander Tans, and Pieter Rein ten Wolde. Transcriptional regulation by competing transcription factor modules. *PLoS computational biology*, 2(12):e164, 2006.

[62] Jorge Trojanowski. Bachelorarbeit, Ludwig Maximilians Universität, 2014.

[63] Michael G. Poirier, Eugene Oh, Hannah S. Tims, and Jonathan Widom. Dynamics and function of compact nucleosome arrays. *Nature Structural & Molecular Biology*, 16(9):938–44, 2009.

[64] K. J. Polach and J. Widom. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *Journal of Molecular Biology*, 258(5):800–12, 1996.

[65] Kien Nguyen. Masterarbeit, Ludwig Maximilians Universität, 2011.

[66] Guo-Cheng Yuan, Yuen-Jong Liu, Michael F. Dion, Michael D. Slack, Lani F. Wu, Steven J. Altschuler, and Oliver J. Rando. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, 309(5734):626–30, 2005.

[67] C. C. Adams and J. L. Workman. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Molecular and cellular biology*, 15(3):1405–21, 1995.

[68] Joanna A. Miller and Jonathan Widom. Collaborative Competition Mechanism for Gene Activation In Vivo. *Molecular and cellular biology*, 23(5):1623–1632, 2003.

[69] Nico Geisel and Ulrich Gerland. Physical limits on cooperative protein-DNA binding and the kinetics of combinatorial transcription regulation. *Biophysical Journal*, 101(7):1569–1579, 2011.

[70] S. Vashee, K. Melcher, W. V. Ding, S. a Johnston, and T. Kodadek. Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. *Current Biology : CB*, 8(8):452–8, 1998.

[71] Bernard Turcotte, Xiao Bei Liang, François Robert, and Nitnipa Soontorngun. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. *FEMS Yeast Research*, 10(1):2–13, 2010.

[72] Najla Gasmi, Pierre-etienne Jacques, Natalia Klimova, Xiao Guo, Alessandra Ricciardi, François Robert, and Bernard Turcotte. The switch from fermentation to respiration in Saccharomyces cerevisiae is regulated by the Ert1 transcriptional activator / repressor. *Genetics*, 198(2):547–60, 2014.

[73] Noam Kaplan, Irene K. Moore, Yvonne Fondufe-Mittendorf, Andrea J. Gossett, Desiree Tillo, Yair Field, Emily M. LeProust, Timothy R. Hughes, Jason D. Lieb, Jonathan Widom, and Eran Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236):362–6, 2009.

[74] Audrey P. Gasch, Paul T. Spellman, Camilla M. Kao, Orna Carmel-Harel, Michael B. Eisen, Gisela Storz, David Botstein, and Patrick O Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, 11(12):4241–4257, 2000.

[75] D. Bourgarel, C. C. Nguyen, and M. Bolotin-Fukuhara. HAP4, the glucose-repressed regulated subunit of the HAP transcriptional complex involved in the fermentation-respiration shift, has a functional homologue in the respiratory yeast Kluyveromyces lactis. *Molecular Microbiology*, 31(4):1205–1215, 1999.

[76] Eva M. Huber, Daniel H. Scharf, Peter Hortschansky, Michael Groll, and Axel A. Brakhage. DNA minor groove sensing and widening by the ccaat-binding complex. *Structure*, 20(10):1757–1768, 2012.

[77] Bertrand Russell. The Philosophy of Logical Atomism. In A. D. Irvine, editor, *Bertrand Russell: Language, knowledge and the world*, chapter 9, page 131. Routledge, London, 1999.

[78] Eran Segal and Jonathan Widom. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*, 19(1):65–71, 2009.

[79] Alexandra B. Lantermann, Tobias Straub, Annelie Strå lfors, Guo-Cheng Yuan, Karl Ekwall, and Philipp Korber. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nature Structural & Molecular Biology*, 17(2):251–7, 2010.

[80] Stephanie Geggier and Alexander Vologodskii. Sequence dependence of DNA bending rigidity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35):15421–6, 2010.

[81] Andrea J. Gossett and Jason D. Lieb. In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae. *PLoS Genetics*, 8(6):e1002771, 2012.

[82] Răzvan V. Chereji and Alexandre V Morozov. Statistical Mechanics of Nucleosomes Constrained by Higher-Order Chromatin Structure. *Journal of Statistical Physics*, 144(2):379–404, 2011.

[83] Răfzvan V. Chereji, Denis Tolkunov, George Locke, and Alexandre V. Morozov. Statistical mechanics of nucleosome ordering by chromatin-structure-induced two-body interactions. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 83(5):5–8, 2011.

[84] Vladimir B. Teif and Karsten Rippe. Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities. *Nucleic acids research*, 37(17):5641–55, 2009.

[85] Kuangyu Yen, Vinesh Vinayachandran, Kiran Batta, R. Thomas Koerber, and B. Franklin Pugh. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, 149(7):1461–73, 2012.

[86] Triantaffyllos Gkikopoulos, Pieta Schofield, Vijender Singh, Marina Pinskaya, Jane Mellor, Michaela Smolle, Jerry L. Workman, Geoffrey J. Barton, and Tom Owen-Hughes. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science*, 333(6050):1758–60, 2011.

[87] Christian J. Wippo, Lars Israel, Shinya Watanabe, Andreas Hochheimer, Craig L. Peterson, and Philipp Korber. The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *The EMBO journal*, 30(7):1277–88, 2011.

[88] J. K. Percus. Equilibrium state of a classical fluid of hard rods in an external field. *Journal of Statistical Physics*, 15(6):505–511, 1976.

[89] J. K. Percus. One-dimensional classical fluid with nearest-neighbor interaction in arbitrary external field. *Journal of Statistical Physics*, 28(1):67–81, 1982.

[90] J. K. Percus. Entropy of a non-uniform one-dimensional fluid. *Journal of Physics: Condensed Matter*, 1(17):2911–2922, 1989.

[91] William Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C : the art of scientific computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

[92] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[93] P. T. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276(1):19–42, 1998.

[94] A. Thåström, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *Journal of molecular biology*, 288(2):213–29, 1999.

[95] Maria Walker, Nils Krietenstein, Johannes Nuebler, Brendan Osberg, Philipp Korber, and Ulrich Gerland. Unpublished data, Experimental data by M.W. and N.K., amoeba inverse software written by B.O., data collected by J.N., 2015.

[96] David M. MacAlpine and Geneviève Almouzni. Chromatin and DNA replication. *Cold Spring Harbor Perspectives in Biology*, 5(8):a010207, 2013.

[97] G. Meersseman, S. Pennings, and E. M. Bradbury. Mobile nucleosomes–a general behavior. *The EMBO Journal*, 11(8):2951–9, 1992.

[98] Andrew Flaus and Tom Owen-Hughes. Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? *Current Opinion in Genetics and Development*, 14(2):165–173, 2004.

[99] Courtney Hodges, Lacramioara Bintu, Lucyna Lubkowska, Mikhail Kashlev, and Carlos Bustamante. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (New York, N.Y.)*, 325(5940):626–8, 2009.

[100] H. Schiessel, J. Widom, R. F. Bruinsma, and W. M. Gelbart. Polymer reptation and nucleosome repositioning. *Physical Review Letters*, 86(19):4414–7, 2001.

[101] Cizhong Jiang and B. Franklin Pugh. A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biology*, 10(10):R109, 2009.

[102] A. Rényi. "On a one-dimensional problem concerning random space filling". *Publ. Math. Inst. Hung. Acad. Sci.*, 4:109–27, 1958.

[103] A. Rényi. On a one-dimensional problem concerning random space filling. *Selected Translations in Mathematical Statistics and Probability*, 4(1):203–218, 1963.

[104] W. Evans. Random and cooperative sequential adsorption. *Reviews of Modern Physics*, 65(4):1281–329, 1993.

[105] G. Tarjus and P. Viot. Statistical mechanical description of the parking-lot model for vibrated granular materials. *Physical Review E*, 69(1):011307, 2004.

[106] Ranjith Padinhateeri and John F. Marko. Nucleosome positioning in a model of active chromatin remodeling enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19):7799–7803, 2011.

[107] Padinhateeri Ranjith, Jie Yan, and John F. Marko. Nucleosome hopping and sliding kinetics determined from dynamics of single chromatin fibers in Xenopus egg extracts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13649–54, 2007.

[108] Jyotsana J. Parmar, John F. Marko, and Ranjith Padinhateeri. Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Research*, 42(1):128–136, 2014.

[109] Brendan Osberg, Johannes Nuebler, Philipp Korber, and Ulrich Gerland. Replication-guided nucleosome packing and nucleosome breathing expedite the formation of dense arrays. *Nucleic Acids Research*, 42(22):13633–45, 2014.

[110] Jairo Rodriguez and Toshio Tsukiyama. ATR-like kinase Mec1 facilitates both chromatin accessibility at DNA replication forks and replication fork progression during replication stress. *Genes and Development*, 27(1):74–86, 2013.

[111] Marta Radman-Livaja, Kitty F. Verzijlbergen, Assaf Weiner, Tibor van Welsem, Nir Friedman, Oliver J. Rando, and Fred van Leeuwen. Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biology*, 9(6):e1001075, 2011.

[112] G. Tarjus and P. Viot. Generalized Car-Parking Problem as a Model for Particle Deposition with Entropy-Activated Rate Process. *Physical Review Letters*, 68(15):2354–7, 1992.

[113] X. Jin, G. Tarjus, and J. Talbot. An adsorption-desorption process on a line: kinetics of the approach to closest packing. *Journal of Physics A: Mathematical and General*, 27(7):L195–L200, 1994.

[114] James B. Knight, Christopher G. Fandrich, Chun Ning Lau, Heinrich M. Jaeger, and Sidney R. Nagel. Density relaxation in a vibrated granular material. *Physical Review E*, 51(5):3957–3963, 1995.

[115] E. Ben-Naim, J. B. Knight, E. R. Nowak, H. M. Jaeger, and S. R. Nagel. Slow relaxation in granular compaction. *Physica D: Nonlinear Phenomena*, 123:380–385, 1998.

[116] J. Talbot, G. Tarjus, and P. Viot. Adsorption-desorption model and its application to vibrated granular materials. *Physical Review. E*, 61(5):5429–38, 2000.

[117] J. P. Sutherland, George Vassilatos, Hiroshi Kubota, and G. L. Osberg. The effect of packing on a fluidized bed. *AIChE Journal*, 9(4):437–441, 1963.

[118] L. Vroman, A. L. Adams, G. C. Fischer, and P. C. Munoz. Interaction of High Molecular Weight Kininogen, Factor XII, and Fibrinogen in Plasma at Interfaces. *Blood*, 55(1):156–159, 1980.

[119] J. L. Brash and P. ten Hove. Effect of Plasma Dilution on Adsorption of Fibrinogen to Solid Surfaces. *Thrombosis and Haemostasis*, 51(3):326–330, 1984.

[120] Amy Kolan, Edmund Nowak, and Alexei Tkachenko. Glassy behavior of the parking lot model. *Physical Review E*, 59(3):3094–3099, 1999.

[121] P. L. Krapivsky and E. Ben-Naim. Collective properties of adsorption-desorption processes. *The Journal of Chemical Physics*, 100(9):6778–82, 1994.

[122] I. Lončarević, Lj. Budinski-Petković, S. Vrhovac, and a. Belić. Adsorption, desorption, and diffusion of k-mers on a one-dimensional lattice. *Physical Review E*, 80(2):021115, 2009.

[123] Galen T. Craven, Alexander V. Popov, and Rigoberto Hernandez. Structure of a tractable stochastic mimic of soft particles. *Soft Matter*, 10(29):5350–61, 2014.

[124] Padinhateeri Ranjith and John F. Marko. Filling of the one-dimensional lattice by k-mers proceeds via fast power-law-like kinetics. *Physical Review E*, 74(4):041602, 2006.

[125] P. L. Krapivsky, S. Redner, and E. Ben-Naim. *A Kinetic View of Statistical Physics.* Cambridge University Press, Cambridge, 2010.

[126] J. Flory. Intramolecular Reaction between Neighboring Substituents of Vinyl Polymers. *Journal of American Chemical Society*, 61:1518–21, 1939.

[127] Erwin Frey and Andrej Vilfan. Anomalous relaxation kinetics of biological lattice ligand binding models. *Chemical Physics*, 284:287–310, 2002.

[128] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).* Wiley-Interscience, 2006.

[129] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier Science & Technology Books, 2007.

[130] Brendan Osberg, Johannes Nuebler, and Ulrich Gerland. Adsorption-desorption kinetics of soft particles. *Physical Review Letters*, 115(8):088301, 2015.